MEASURING THE EFFECTS OF DISPLAY DESIGN AND INDIVIDUAL DIFFERENCES ON THE UTILIZATION OF MULTI-STREAM SONIFICATIONS

A Dissertation
Presented to
The Academic Faculty

by

Jonathan H. Schuett

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Psychology

Georgia Institute of Technology

August 2019

MEASURING THE EFFECTS OF DISPLAY DESIGN AND INDIVIDUAL DIFFERENCES ON THE UTILIZATION OF MULTI-STREAM SONIFICATIONS

Approved by:

Dr. Bruce Walker, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Rick Thomas
School of Psychology
Georgia Institute of Technology

Dr. Frank Durso
School of Psychology
Georgia Institute of Technology

Dr. Michael Nees
Department of Psychology
Lafayette College

Dr. Richard Catrambone
School of Psychology
Georgia Institute of Technology

Date Approved: [April 29, 2019]

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Bruce Walker, for his many years of guidance and mentorship. I would like to thank the School of Psychology faculty for challenging me and pushing me to do great things. And I would like to thank the many members of the Sonification Lab for their support and friendship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Previous work in the auditory display community has discussed the impact of both display design and individual listener differences on how successfully listeners can use a sonification. This dissertation extends past findings and explores the effects of display and individual differences on listeners' ability to utilize a sonification for an analytical listening task when multiple variables are presented simultaneously. This is considered a more complicated task and pushes listeners' perceptual abilities, but is necessary when wanting to use sonifications to display more detailed information about a dataset. The study used a two by two between-subjects approach to measure the effects of display design and domain mapping. Acoustic parameters were assigned to either the weather or the health domain, and these mappings were either created by an expert sound designer or arbitrarily assigned. The acoustic parameters were originally selected for the weather domain, so those display conditions were expected to result in higher listener accuracy.

Results showed that the expert mapped weather sonification led to higher mean listener accuracy than the arbitrarily mapped health display when listeners did not have time to practice, however with less than an hour of practice the significant main effects of design and domain mapping went away and mean accuracy scores increased to a similar level. This dissertation introduces two models for predicting listener accuracy scores, the first model uses musical sophistication and self-reported motivation scores to predict listener accuracy on the task before practice. The second model uses musical sophistication, self-reported motivation, and listening discrimination scores to predict listener accuracy on the sonification task after practice.

CHAPTER 1. INTRODUCTION

It is not overly difficult to imagine scenarios where users want to receive information through audio. As more of the devices we routinely interact with become constantly connected to data sources, but screen real estate remains somewhat finite, presenting users with information through speakers or headphones will continue to be a viable option. Whether it is a visually impaired student wanting to learn statistics by listening to representations of graphs, an investor wanting to check their stock portfolio while driving to work, or a sleepy-eyed user wanting to listen to the daily weather forecasts while getting out of bed in the morning, there is a large amount of information that can be passed from auditory displays to the user.

A determining factor in how useful an auditory display is for a given scenario may depend on how well the designer understands the abilities and limitations of listeners' auditory perceptual system. One of these limitations is often discussed as, *how much information can we put into an auditory display?* (Flowers, 2005; Nees & Walker, 2007; Schuett, Winton, Batterman, & Walker, 2014). Perhaps stated more formally, *how many concurrent sounds can someone listen to at a given time?* While this question seems somewhat straightforward, it is an ongoing subject of debate among the auditory scene analysis community as well as a continued topic of discussion within the auditory display research community. The seemingly simple question starts to become complex when we begin to think about it in more specific ways. For instance, the number of *sounds* may refer to discrete acoustic events or an ongoing auditory soundscape. Additionally, listeners may only need to perceive that a number of sounds are present, but often they are required to attend and comprehend the meaning of stimuli that they receive.

1.1 Background

This document will discuss the background and overview of both sonification design and the perceptual process of auditory stream segregation. After establishing the connection between these two research areas I will begin to discuss the factors that impact stream segregation during listeners' perception of auditory displays. The investigation includes factors of display design, training, and task type. Display design includes all acoustic aspects of how a sonification or auditory display is created. Training will discuss any aspect of experience and feedback that listeners receive. And task type will discuss the different ways in which listeners may be expected to interact with a display.

1.1.1 Sonification and Auditory Display

While it may seem like the terms *auditory display* and *sonification* are being used interchangeably through this document, it is important to understand the distinction between the two. An auditory display refers to any sound-based interaction between a human user and a technology interface (Kramer, 1994b). A sonification is slightly more specific in that it is "the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation" (p. 4) as defined by the Sonification Report (Kramer, Walker, Bonebright, Cook, Flowers, Miner, & Neuhoff, 2010; also Scaletti, 1994). Sonification is a form of auditory display with a specific focus on conveying the relationships between data. The term auditory display is used in this document to refer to a more encompassing group of audio-based displays in general, with the primary focus on sonification design.

1.1.2 The Auditory Stream

To help create a common terminology among researchers, Bregman and Campbell (1971) suggested the term *auditory stream* as a way of discussing brief auditory occurrences. This is because multiple acoustic events can blend together and the listener can perceive them as a unified idea or stream. This prevents us from having to resort to the unhelpful task of breaking down and differentiating between every individual acoustic event, such as each sound an individual rain drop makes among thousands of rain drops, versus the overall auditory concept for the sound of a rain storm. Discussing sound as a stream is perhaps the equivalent of referring to the conceptual sound of rain rather than thinking of it as thousands of drops of water hitting the ground, trees, and buildings around you. When considering an auditory stream, it can be thought of as something that acts as a whole, but also consists of smaller parts. Similar to the discussion of elements, molecules, and atoms.

The concept of an auditory stream becomes useful for how we discuss the perceptual process that listeners use to group or segregate sounds within their listening environment. Listeners use this process to create an understanding or awareness of different sound sources and some level of *what* and *where* things are around them. This is the primary focus in Al Bregman's book, titled *Auditory Scene Analysis* (1990), which goes into great detail on how auditory streams form from a sometimes overwhelmingly vast number of auditory events, and how listeners use and rely on the human ability to parse concurrent auditory streams from one another to make sense of an acoustic environment. While scene analysis has its obvious ecological uses for listeners, the concept at this point was primarily discussed in terms of better understanding how or maybe even why we listen to sounds in the world around us. There was little discussion about using scene analysis to explain how listeners interact with designed sounds artificially placed into the environment to convey information streams to listeners.

In 1994, Greg Kramer published the book *Auditory Display*, which acts as a compendium for the proceedings and discussion from the first International Conference on Auditory Display [ICAD] (Kramer, 1994a). This group of researchers and sound designers met to discuss the uses, benefits, and implications of using non-speech audio to convey information to listeners. Al Bregman – the same who wrote Auditory Scene Analysis – wrote the foreword for Kramer's book. While Bregman's work has focused more on theories of psychoacoustics rather than applied display design, the connection between scene analysis and auditory display research emerged. In his foreword, Bregman admits that he had not previously considered the major benefits of non-verbal auditory displays over verbal or visual displays for monitoring tasks, comparing large real-time datasets, or multitasking between visual and auditory tasks. In his words, the decision to use auditory displays for specific tasks become clear; "use each sensory modality for what it does best."

1.1.3 The Formation and Segregation of Auditory Streams

Even before the term *auditory stream* (Bregman & Campbell, 1971), psychoacousticians and linguists were interested in the point at which sounds were either perceptually grouped or segregated from one another. Early relationships between factors such as the presentation rate of a series of tones and their frequencies were found to have an effect on whether the tones would be perceived by listeners as a single stream or multiple overlapping streams (Heise & Miller, 1951; Miller & Heise, 1950; van Noorden, 1975). Miller and Heise (1950; Heise & Miller, 1951) discovered a phenomenon that they called the "Trill Threshold" which is based on the perceived relationship between two repeating tones. With one of these tones [A] at a higher frequency than the other [B], if they are presented in a repeated pattern [A-B-A-B-A-B-A…] the listener will report hearing either one stream that "gallops" between the two tones [ABABABA] or they may report two streams playing simultaneously. These two streams may seem like a higher frequency tone [AAAAAAA] and a lower frequency tone [BBBBBBB] played in unison. Whether listeners reported hearing one stream or two streams seemed to rely mostly on how much frequency separation there was between the two tones. Miller and Heise altered the frequency separation between tones A and B until a threshold could be established for the point at which the tones perceptually formed separate streams. As the presentation rate between the two alternating tones increased (i.e., less time between tones), the less frequency separation between the tones needed to be for stream segregation to occur. This indicates an inverse relationship between frequency separation and presentation rate of tones and their likelihood to segregate into separate streams.

Holding the frequency of one tone [B] constant and using the method of adjustment to allow the other tone [A] to increase or decrease in frequency, van Noorden (1975) was able to replicate earlier findings of the Trill Threshold tradeoff of presentation rate and frequency, while presenting the tones already grouped in patterns of ABA-ABA-ABA-ABA… While the researcher altered the presentation rate of these tone groups, the listener would alter the frequency of tone A until A and B either fused into one stream or segregated into two. Van Noorden described this fusing of tones as "Temporal Coherence," while the segregation into separate streams he called "Fission." A depiction of these occurrences and the relationship with frequency can be seen in Figure 1.

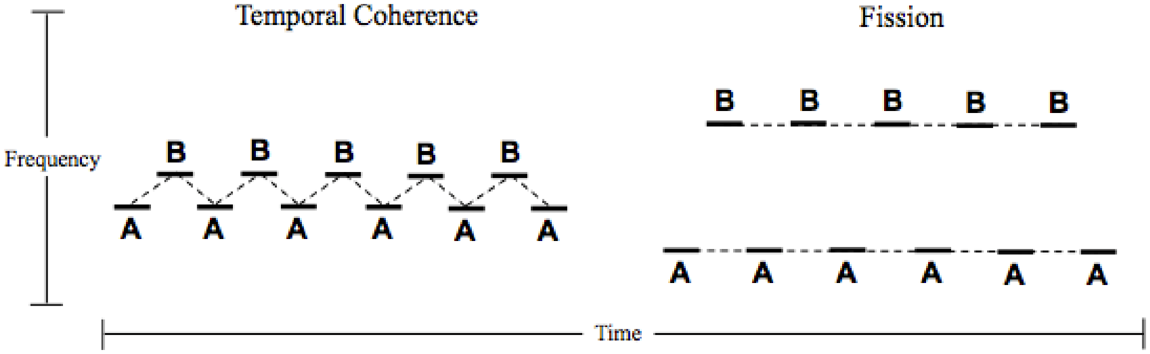


Figure 1. Temporal Coherence [ABABABA] and Fission [AAAA + BBBB] as described by van Noorden (1975).

1.1.4 Mechanisms & Theories for Why Stream Segregation Occurs

With evidence for Temporal Coherence and Fission, van Noorden (1975) attempted to explain this perceptual phenomenon as the result of rapid triggering of overlapping hair cells in the ear. More specifically, van Noorden was likely talking about the cilia (hair cells) located along the Basilar membrane in the cochlea. These cilia trigger in different locations within the cochlea specific to tone frequencies as the listener senses acoustic stimuli. In general, van Noorden's theory seems to be based on a refractory period for the firing of the nerve cells which have a physical limit to how often they can trigger. Van Noorden believed when back-to-back tones of the same or similar frequency cause excitation of the cilia along the cochlea at the same time or within rapid succession, the tones would not be perceptually different enough, thus resulting in the perception of only one stream (Temporal Coherence).

While this explanation seemed plausible for van Noorden's early findings, we know it is not robust enough to account for things like backwards masking (Massaro, 1972), nor findings that complex tones (tones with multiple harmonic frequencies) containing the same fundamental [overlapping] frequency can still segregate into different streams (Bregman & Liao, 1984; Bregman, Liao, & Levitan, 1990). Considering at least some of these other findings, van Noorden offered a slightly more comprehensive explanation suggesting that perhaps some physiological structure was accounting for the changes in frequency in addition to the hair cells in the ear. This is very similar to the later explanation offered by Anstis and Saida (1985), which suggested *change detectors* that specifically pick up when the frequency change occurs. They theorized that change detectors could not keep up when presentation rate of tones was too rapid, leading instead to the perception of multiple streams.

Neither of these two explanations account for any level of listener control over stream segregation. Norman (1967) used a series of target tones and background tones (distractors), asking listeners to judge where the target probe tone fell within the acoustic sequence. Based on his findings he concluded that the listener must shift their "attention" between the different frequency tones as orders in the sequence differed. From this it seems Norman was offering a more top-down explanation, implying that it is the listener who must pick out the target stream from the background stream. However, this explanation relies too much on listeners actively trying to follow streams, while most of the previously discussed studies were finding that stream segregation occurred without conscious effort.

Disagreeing with both van Noorden's explanation of a physiological filter and Normans's attention centered approach, Idson and Massaro (1976) suggest that there must be some sort of structural process where the nervous system plays a critical computational role in early processing stages. They believe that frequency differences between adjacent tones in sequences at slower presentation rates could be streamed together in temporal coherence, but when presentation rate increases to a point at which the physiological structure is overwhelmed, Fission will occur and the listener perceives two separate auditory streams due to a breakdown in the processing of tone integration.

Bregman (1990) does not support these theories and points out a few of their flaws. An issue with any explanation suggesting that stream segregation occurs due to over excitation of similar frequency bands on the cochlea or auditory nerves fails to account for streaming that occurs between tones that are presented simultaneously into different ears in dichotomous listening tasks (Deutsch, 1974; Deutsch, 1975; Treisman, 1971). To extend the frequency-based explanation while accounting for evidence of dichotic stream segregation, Hartmann and Johnson (1991) suggest a tradeoff between two peripheral *perceptual channels*. The first being frequency based and reliant on signals from the cochlea, which they described as the *tonotopic* channel. The second channel they called the *lateral* channel and relies on which ear presentation occurred in.

Dichotic listening tasks, or tasks that present separate stimuli to both ears, often require the listener to either utilize or ignore information presented into opposite ears. Much of these tasks pertain to trying to understand the “cocktail party problem” (Cherry, 1953; Yost, 1997), or the ability of a listener to attend one stream while occasionally shifting to information in another. Often these stimuli were verbal based sentences (Cherry, 1953; Treisman, 1960; 1964; 1971), but stream segregation and temporal coherence have been found to occur between non-speech acoustic stimuli presented in different ears as well

(Deutsch, 1974; 1975). Hartmann and Johnson (1991) suggested that peripheral channeling accounts for most of the stream segregation effect. While this claim has been supported by more recent research (Beauvois & Meddis, 1996; McCabe & Denham, 1997; Rose & Moore, 2000), there is still admittedly a gap in this theory for explaining how stream segregation occurs for stimuli characteristics that are less physiological and more perceptual in nature (Moore & Gockel, 2002).

One of these perceptual acoustic characteristics is timbre. Timbre is often defined as the uniqueness of a sound while holding all other acoustic features (frequency, intensity, etc.) constant (Houtsma, 1997; Wessel, 1979). Timbre is perhaps best realized through the example that it is the perceptual quality that makes a violin sound like a violin and a trumpet sound like a trumpet if both instruments were playing the same note (frequency) at the same loudness (intensity). While this perceptual difference is likely influenced by the shape and physical features of the sound source, such as the resonance of the violin's string vibrations through its hollow wooden body versus the metal tubular structure of the trumpet, this physical explanation does not account for the experience of perceiving a trumpet when the note is created artificially on a computer and played through speakers. While the relationship between sound source and timbre is still somewhat ambiguous, multidimensional scaling work in psychoacoustics has shown a relationship between acoustic parameters, such as brightness and the spectral envelope of a sound wave, as impacting listeners' perception of timbre (Grey, 1977; Grey & Moorer, 1977; Gordon & Grey, 1978; McAdams, Beauchamp, & Meneguzzi, 1999; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995; Wessel, 1979). Although the impact of these physical attributes of a sound wave can be scaled, timbre itself as the listener experiences it, is considered to be completely perceptual (Wessel, 1979) and sometimes even subjectively related to frequency (Houtsma, 1997).

Timbre has been found to encourage stream segregation (Bregman & Pinker, 1978; Hirsh, 1959; Schuett, 2010; Singh, 1987; Warren, Obusek, Farmer, & Warren, 1969) often leading to a possible tradeoff between frequency and timbre as they compete for which is the stronger grouping cue for the segregation of tones into multiple streams (Schuett, 2010; Singh, 1987; Singh & Bregman, 1997). Singh (1987) presented participants with tone sequences that altered in frequency and timbre at various presentation speeds. Participants were asked to select from images used to represent patterns in the acoustic stimuli. Height changes in the y-axis of objects in these images represented frequency changes. Objects were depicted with circles or stars to represent different timbres. Listeners were able to utilize both timbre and pitch (frequency) to perceptually organize tones into streams.

Considering the tradeoff between frequency and timbre on how acoustic elements are encouraged to segregate or stream together, an explanation for stream segregation needs to account for these perceptual based factors in addition to the physiological ones. Cusack and Roberts (2000) presented listeners with a series of tones under a variety of conditions that differed by timbre, frequency region, or both. Confirming the occurrence of stream segregation between tones when frequency differences were not present, Cusack and Roberts concluded that peripheral channeling (Hartmann & Johnson, 1991; Moore & Gockel, 2002; Rose & Moore, 2000) could not account for these findings.

Instead, Cusack and Roberts offer an alternative explanation, suggesting that both bottom-up and top-down processes are occurring. The bottom-up process is occurring as “primitive stream segregation” likely taking place somewhere between excitation on the cochlea and its subsequent neural pathway. This is where the frequency range and presentation rate tradeoff impacts stream segregation, similar to what Idson and Massaro (1976) had suggested. In addition to this, Cusack and Roberts discuss a top-down “Schema-Driven Selection” (Bregman, 1990) that accounts for perceptual grouping of items into streams. This Scheme-Driven Selection accounts for perceptual characteristics, contextual factors, and the listener’s attention. If both are always working, top-down selection can override the incoming primitive segregation, or the two can work together to encourage additional stream segregation through redundancies.

1.1.5 Boundaries on the Number of Auditory Streams

The stream segregation studies discussed up to this point have used stimuli that rely on pure tones, which are created artificially to provide control over the random variations found in more complex tones that occur in natural environments. For instance, when a single string in a piano is played (a key is pressed and a hammer strikes the corresponding string) other strings around that which was struck also resonate creating a more complex sound wave. To examine listener limitations in a more ecologically valid scenario, Huron (1989) presented listeners with Bach’s Fugue in E-flat as his stimulus. This piece contains a variation between one to five musical *voices* that change throughout. Treating these voices as auditory streams, Huron asked participants to identify how many piano-voices they could hear changing throughout the piece. Huron found that when more than three voices [streams] were present in the music, listeners would become perceptually confused and report a lower number. He concluded that listeners seem limited to three perceptual auditory streams, but admitted that perhaps timbre differences could lead to more streams considering he only used one instrument.

A later study (Schuett, 2010) used a similar method by providing listeners with up to five different unique tonal elements differing in frequency and timbre. Timbres for this study were selected for greatest differences in spectral envelope and brightness in order to create an optimal combination of frequency and timbre differences that could encourage segregation of up to five streams. Altering the presentation rate of tones and including piano-only and sawtooth-wave-only sets of tones as controls, participants were still only able to report hearing three streams regardless of the timbres used as presentation rate increased. Under these optimal conditions for the segregation of as many auditory streams possible, listeners tend to be limited to about three streams when played simultaneously or at fast presentation rates.

1.1.6 Application in Auditory Display

Wanting to apply a more methodological approach to the design of auditory displays, and interested in the number of tones [differing in frequency] that a listener can identify within a judgment task, Pollack (1952) suggested a methodology based on the information theory approach (Pollack 1953a; Shannon, 1948). He determined that if it is possible to quantify the amount of information presented to the listener, while also finding a way to quantify the amount of information that is lost by the listener, then we can measure how much information is received, gained, or transferred.

Presenting listeners with a series of tones that differed only by frequency, participants were instructed to listen to a 2.5 second tone, assign a numerical label to it, and then listen for the next tone after a 25 second delay. For each successive tone, listeners would assign a new number if it were novel, or recall its numerical label if it was a previously used tone (repeat frequency). The initial findings of this study show that as the amount of transferred information increases, the amount of information gained or information that is useable by the listener tops out around perfect identification for only four tones [unique frequencies]. With several days of practice, this upper limit increases to about 5 tones. Pollack suggested these findings may have been limited due to the one-dimensionality of his display design, proposing that a two- or three-dimensional display would likely result in a higher number of identifiable tones. We should note that he is referring specifically to acoustic dimensions.

Exploring this further, Pollack presented the tones in a similar frequency-based design with the addition of intensity changes (perceived by the listener as loudness). This created the effect of a two-dimensional display, and later a third dimension was added, which contained an anchor or comparison tone. Pollack (1953b) found slight improvements as the number of dimensions increased within the stimuli. The number of usable tones could possibly increase if he had used a parameter like timbre, which encourages a greater amount of perceptual streaming than intensity differences (Singh, 1987).

1.2 The Schematics of an Auditory Display

Kramer (1994b, see also Walker & Nees, 2011) describes an auditory display system in three parts. Starting with the *Information Generator*, which for most purposes is some form of database that the display pulls from, is the source of the data being displayed. Next is the *Communicative Medium*, which consists of the processes that must occur by the auditory display system. This includes retrieving the data, processing the data, and then generating sounds for the display. The final piece of this system is the *Information Receiver*, which is usually the human listener. This information receiver must sense, perceive, and process the sounds in order to understand the display (Walker & Kramer, 2004). Figure 2 shows these three parts of the system in what Kramer (1994b) titled the "Schematic of an Auditory Display System." This schematic could be updated to simplify the three parts into, the *Data Source*, the *Auditory Display* (and underlying mechanisms), and the *Listener*. Both versions are included in Figure 2, the simplified version is below Kramer's original model.

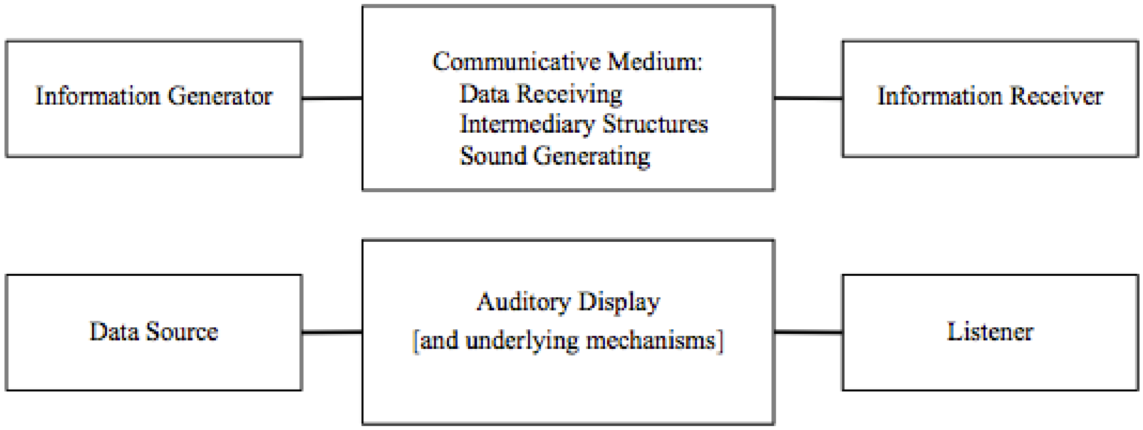


Figure 2. Kramer's (1994b) Schematic of an Auditory Display System along with an updated version for simplicity.

Rather than focus on the relationship between the data source and its impact on how the data is displayed, this dissertation focuses on the relationship between the listener and the display. Instead of treating the lines on Kramer's model as one-way arrows, this should instead be an iterative process. It is important to realize that the relationship between the listener and the data type is necessary for determining how the auditory display should be designed and used. Figure 3 suggests a more accurate representation between these relationships.

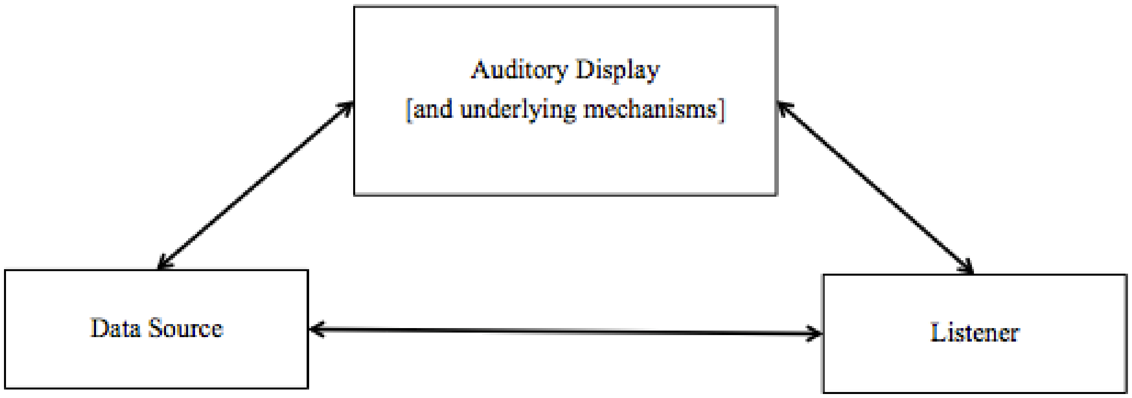


Figure 3. The iterative process and relationship between the data sources, auditory display, and listener.

1.3 Sonification Design Example

Building a sonification of stock market data, the simplest sonification could utilize one auditory stream. For this display, the value of a stock could be mapped to a continuous tone that goes up or down in frequency (which is perceived as musical pitch) as the value of the stock rises or falls through a given period of time (hours, days, weeks, months, etc.). But how useful is this single stream for the listener? The listener can monitor the direction their stock is moving, for instance if the pitch is in a continuous decline, the listener may want to sell their stock before it loses too much value. But with this single stream sonification the listener is limited to a monitoring task, and does not have any other context aside from the direction in which the data is changing and the rate at which this change is occurring (Nees & Walker, 2007; Smith & Walker, 2002; Smith & Walker, 2005). Think of this simplest sonification example as the equivalent of a line drawn on a piece of paper, as depicted in Figure 4.

Figure 4. Single line representing the value price for shares of stock.

To make this simple trend line more informative we can add additional context. This can be applied to the sonification with additional streams of information (Smith & Walker, 2002; 2005). For instance, we draw x- and y-axis lines with tick marks to provide greater visual detail about how much time has passed [x-axis] and how many value points the stock changes by [y-axis]. This can be seen in Figure 5.

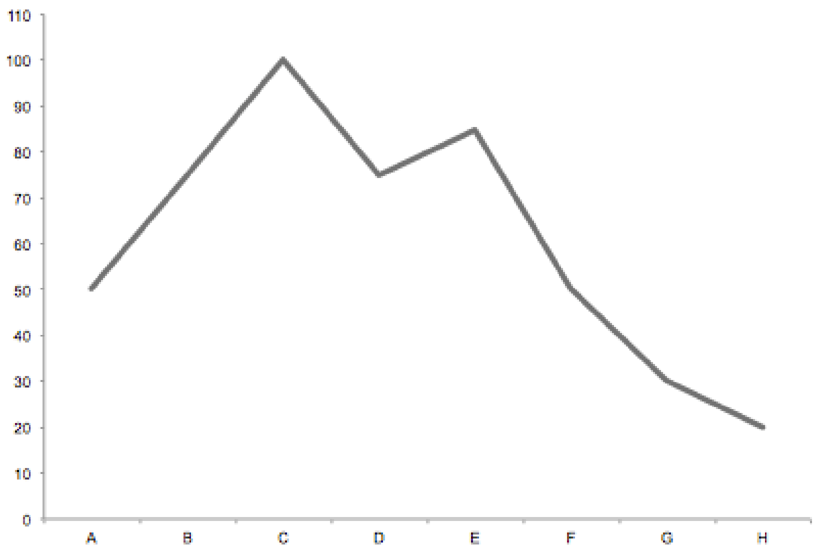


Figure 5. Stock data and context represented visually with values on the x- and y-axes.

In the auditory variant of this display, the tick marks on the x-axis could be represented with the sound of a bell that occurs when a unit of time has passed. Similarly, the y-axis tick marks can be represented with a whistle each time the value of the stock crosses another value point. The example sonification has the potential for three different auditory streams: a continuous tone changing in pitch over time, the discrete bell tone for the tick marks signaling the passing of time on the x-axis, and the discrete whistle sound for tick marks each time a new point is crossed on the y-axis. This sonification design example could continue to progress until it becomes almost unnecessarily complex, containing axes labels, alerts on specific thresholds set by the user for things like when a stock value goes above or below a set amount, and could even include a second dataset for comparison purposes. A visual version of this is depicted in Figure 6.

The graph displayed in Figure 6 is somewhat over complicated for the purposes of this example, yet the use case of monitoring stock values would easily require much of the discussed information in order to be fully useful. Not all of that data may need to be displayed simultaneously, but in an accessible display for visually impaired users who want to monitor their stock, some form of sonification is needed for each of these information streams.

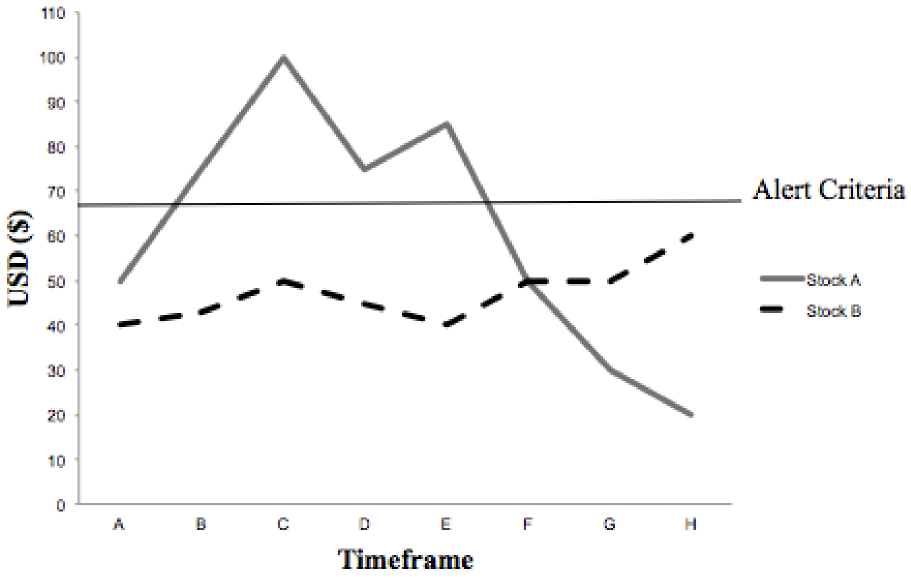


Figure 6. Complex visual graph of stock data example, including axes labels, comparison stock, and alert criteria for specified values.

A sonification can be designed so that more than one type of data is represented in a single auditory stream. Considering our stock market data example, the x-axis data of time may not need to be a separate stream but could be represented by building it into the initial stock quote stream (the one changing in frequency). Rather than a continuous tone that rises or falls in frequency over time, the tone could be represented as discrete tones, which still change in pitch, but each ‘beat’ could represent the unit of time being displayed across the x-axis. A visual depiction of this is shown in Figure 7.

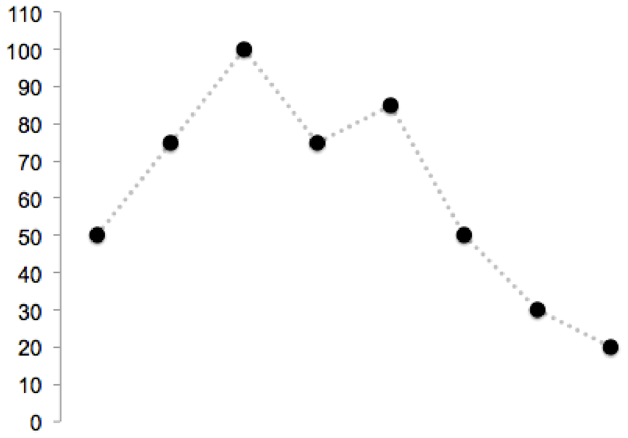


Figure 7. Discrete tones over time for stock quite with value still represented along the y-axis, but for this example instead of tick marks for the x-axis a discrete tone is played representing each time point.

The overall shape of the data is intact and the listener can still follow the trend of the data values, represented as frequency over time. If the listener knows what unit of time is being represented across the x-axis, then each beat of the tone can represent that unit along what would have been the tick marks (Smith & Walker, 2002). The y-axis could still be represented with the whistle tone when a value range has changed across a threshold, as it was in the previous example. With this example we now have a useful sonification with multiple information sets across two auditory streams. One stream contains two types of information, one changing in frequency and the other represented by the beat signifying passage of time on the x-axis. The other stream has one information type dedicated to representing the value of stock with occasional whistle tones as the values change between place markers.

1.4 Display Design

Brewster, Wright, and Edwards (1994) provide guidelines of acoustic parameters for the design of sonifications in the form of ‘earcons’ or short abstract auditory presentations used to provide listeners with information. Barrass (1997) turned these guidelines into a table to be used as reference when designing auditory displays, discussing acoustic and musical attributes to be used in sonifications. Kramer (1994b) compares the contributions of composers towards auditory displays as analogous to the impact that graphic designers have on visual displays.

Many of the acoustic principles discussed for auditory display design are the parameters that relate to stream segregation. For instance, timbre, is a perceptual quality that can lead to the segregation between two streams of auditory elements. Therefore, it makes sense that Brewster, Wright, and Edwards (1994) suggest using distinctly different timbres to make separate earcons stand out from one another (i.e., form separate streams). The same can be said about pitch, measured physically as frequency, which is one of the primary cues to encourage auditory streams to segregate or group. It is no surprise that Brewster et al. suggest using large frequency separation between tones used for earcons so that they create their own streams.

1.4.1 Mapping

Although it would not be difficult to design a display by arbitrarily selecting the acoustic properties that seem to sound best, Walker and Kramer (1996) have shown that it is important to understand the relationship between the acoustic dimensions and how they map to data dimensions within a sonification. Walker and Kramer created sonifications of an imaginary factory and asked participants to listen for changes. Mappings based on stronger or natural metaphors resulted in better listener performance on the monitoring task. Walker (2000; 2002) presented listeners with a variety of acoustic parameters that mapped to variables such as temperature, pressure, velocity, and size. Using the psychophysical technique of magnitude estimation (Stevens, 1957; 1975), Walker would present a tone and ask participants to assign a numerical value for what they felt it represented in relation to the specified physical variable (e.g., temperature). The participant would then listen to a second tone, changing by some magnitude along one of the acoustic dimensions (e.g. pitch/frequency). Next the participant would assign a value for what they thought this new tone represented for the variable of interest. Over many trials and comparisons between different levels of each acoustic parameter and each concept mapping, Walker was able to calculate how different levels of change on acoustic dimensions like pitch (frequency) or loudness (intensity) relate to how much change listeners perceive to have occurred along the data. For example, a 40 Hz change in frequency may represent a 5 degree change in temperature.

Walker found that some acoustic variables map better to specific data types, for instance temperature changes map well to frequency, but not to intensity. He also found that the polarity and scaling were important to consider. Polarity refers to the direction in which conceptual changes in data map to changes in the acoustic property. Scaling is the consideration of how large an acoustic change should be in relation to the data that is being represented. Walker and Mauney (2010) used magnitude estimation to evaluate polarity mapping for frequency, tempo, and acoustic modulations. They discovered a reverse polarity for participants who were visually impaired compared to responses from sighted participants. Sighted participants mapped values like ‘amount of money’ as a positive polarity, so that an increase in pitch (frequency) was equal to an increase in monetary amount. Blind participants tended to report an opposite polarity mapping. Concepts like size or velocity offer physical experience, but conceptual values like dollar amount may need to be thought about more abstractly. Sighted listeners may be more familiar with visual graphs representing monetary amounts and may have used that as their default. However blind participants who have never seen a visual graph have no mental model for how one should be displayed. Walker and Mauney concluded that the visually impaired participants must rely on other experiences when trying to determine how the polarities map. This hypothesis was supported anecdotally when one of their blind participants explained that she thought of decreases in pitch as an increase in money because having a bigger bag of money would be heavy, and heavy things tend to have a deeper sound. The findings of this and similar studies imply that mapping and polarity choices for representing data in a sonification may need to be based on the data type (Walker, 2000; 2002), the user population (Walker & Lane, 2001), or even listener preferences (Walker, Kramer, & Lane, 2000).

1.4.2 Interaction Between Acoustic Parameters and Judgments

An issue to consider when determining how data should be mapped to acoustic parameters is how the interaction between multiple acoustic parameters may alter listeners' judgments. For instance, Melara and Marks (1990) measured participant reaction times for judgment tasks and found that participants responded faster to tones with a "hollow" timbre in higher frequency ranges, tones with a "twangy" timbre were responded to faster at lower frequency. Changes in frequency are found to influence judgments of loudness (intensity), as well as the reverse, with loudness influencing pitch judgments (Neuhoff & McBeath, 1996; Neuhoff, McBeath, & Wanzie, 1999). When listeners hear both pitch and loudness changing in the same direction, they tend to report greater magnitude for both (Neuhoff, Waynard, & Kramer, 2002). Even changes within a single parameter can alter judgments; Neuhoff et al. (2002) found that rising intensity of sounds were perceived as representing greater amounts of change than falling intensities. Using selective-listening tasks for both pitch and pitch-change stimuli, Walker and Ehrenstein (2000) found that participants' responses were faster for pitch than for pitch change and determined that listeners cannot completely isolate just one attribute within a sonification. Bly (1994) points out that one of the problems with audio representation of multivariate data is that "acoustic parameters are not independent, nor equally prominent" (p.406). McGookin and Brewster (2004) found that listeners struggled to correctly identify overlapping earcons within a display, but when different timbres were added to the earcons, identification performance improved. Judgment tasks as far back as 1932 (Chapman, 1932) have shown that accuracy of judgments decreases as the number of acoustic dimensions increases. Therefore, it is important to be mindful of how many attributes are included and how those acoustic attributes will impact how data is grouped into streams.

1.4.3 Mapping to Streams

Kramer (1994c) claims that if a listener is spending more resources trying to remember how data is mapped to different auditory parameters, they may not focus enough of their attention on the trends in the data itself. To avoid this, Kramer recommends representing common data types with common auditory parameters. These representations are called "data family / stream associations." Considering the auditory scene created by these sonification streams, Kramer discussed the relationships between data and how perceptual grouping would affect comprehending changes in the data families themselves. The following associations are what he determined to be potential occurrences between data families.

PER STREAM. In a *per stream* association between the data and the acoustic variable, the acoustic variable represents only one type of data family. For example, if we consider a sonification of population densities for different types of birds on campus, Finches may be represented by the timbre of a violin. The timbre of a flute could be assigned to represent sparrows. The *per stream* example here is that of each unique timbre type to a set data variable type. Each instrument timbre can be recognized by the listener as a unique sound source so it is likely that each would form its own stream. In this case, the listener may attend to information within one stream or the other. While listening to the sonification of both populations simultaneously, you may focus only on the violin timbre if you are interested in just the finch population.

INTER-STREAM. In this type of association, one variable may affect more than one data family, but not all of the data families in the sonification. If we return to our bird populations on campus example, some birds may be more seasonal than others. For example, if our sonification included population density for finches, sparrows, and grosbeaks, time of year would impact population size but only for those that migrate. If only sparrows and grosbeaks are migratory, then time of year affects those data families greater. Kramer suggests using an *inter-stream* association for data that affects some but not all of the data families. Perhaps in our example lowering the intensity, or loudness, could be used to signify when the migratory bird data families are affected by time of year. With this type of *inter-stream* mapping, the listener can still listen to just one stream to get information about the population of a specific species, and know if and when that population migrates.

GLOBAL. When a data variable affects all of the data families, an acoustic parameter that can affect all of the streams within the sonification should be used. If there was a time reflected in the data where there was also a population of feral cats living on campus this could affect all bird populations in the area, this inverse relationship could be represented across all of the auditory streams. For instance, we may adjust the rhythm or pattern of beats within the sonifications, affecting how all three of the bird populations are represented during the time at which the feral cat population on campus became unmanageably high and resulted in above average amounts of birds being eaten. As the feral cat problem is cleared up and bird populations are no longer affected, the rhythm of each stream returns to normal. For this type of association in the data, the listener can attend to the entire auditory scene for all of the streams together, if they are interested in the overall effect of the feral cats on bird populations as a whole. Or the listener still has the option to focus on a single stream if they are interested in just one of the species at a time.

While Kramer's suggestions for how data families and associations across acoustic parameters attempt to address how sonification users can attend data associations between streams, Barrass (1997) explains that separating information between streams can make tasks more difficult for the listener. He provides us with these three consequences (p. 78):

- Streams are categorical and exclusive
- Judgments involving elements in the same stream are easy
- Judgments involving elements in different streams are difficult

Barrass explains these outcomes in terms of the Primitive Stream Segregation and Schema-Driven Selection explanations. He describes the *global* associations that Kramer (1994c) talks about between-streams as part of that primitive process, while the local associations achieved in *per stream* and *inter-stream* variables are attended by the listener and are top-down controlled, fitting the Schema-Driven stream selection process. Making judgments about elements between streams requires the shifting of attention and requires more resources of the listener.

1.4.4 Sonification Configuration

Both Kramer (1994c) and Barrass (1997) discuss the consequences of distributing information from different data variables across streams. An alternative approach looks instead to load multiple data variables onto fewer streams. This allows a designer to plan for the listener to attend as few streams as possible while still receiving information from multiple variables. This intentional design with an intended number of perceptual streams used to carry multiple pieces of information is called *sonification configuration* (Anderson & Sanderson, 2004). Using this technique, Fitch and Kramer (1994) found that not only could listeners successfully monitor eight different variables across two different streams, but they could perform better using a sonification compared to a visual depiction or an audio/visual combination. Participants took on the role of anesthesiologist and were given the task of keeping their 'digital patient' alive by monitoring vitals. During this monitoring task, changes could occur on one, two, or three different variables at any given time. Loeb and Fitch (2002) replicated this study using actual anesthesiology residents and six variables representing vital signs across two acoustic streams. While participants did not show an increase in accuracy rate for detection nor identification of the variables, they did show a significant decrease in response time averages for the condition that combined a visual display with the 2-stream sonification.

This anesthesiology sonification task was extended by Anderson and Sanderson (2004; 2009) during which, participants were presented six variables of vital signs represented in sonification configuration mappings of one, two, or three streams. In the one-stream condition all six variables were ‘carried’ together, on the two-stream condition three variables were ‘carried’ on each stream, and on the three-stream condition two variables were on each stream. When one variable changed during the task, participants were most accurate when monitoring one stream. But when multiple variables changed simultaneously, participants were more accurate at detecting the changes if the variables were dispersed across more streams.

1.5 Individual Differences

Watson and Kidd (1994) propose that humans have perceptual and decision-making skills that are worth “honing to a fine edge.” That is to say, much of the listener’s role in an auditory display system is based on perceptual and cognitive abilities, and how well the system performs can be impacted by the listener’s ability. Cusack, Decks, Aikman, and Carlyon (2004) point out that when listeners know what they are listening for, they can selectively attend specific acoustic characteristics of the target sound or sound source and devote more cognitive resources to that stream. Dowling (1973) found that participants were more accurate at identifying a target melody when they were familiar with the target. However, experience with the background streams did not aid in identification of the target. This demonstrates that knowing what the target is may in fact allow the listener to focus attention on similar acoustic parameters, streaming them together, but knowing what distractors to ignore does not necessarily help segregate the target stream from distractor streams any better.

Findings like these lead to some confusion on what type of training or experience is helpful for the listener. Kidd and Watson (1987) presented listeners with a series of complex tones that differed by three acoustic parameters. Listeners were asked to discriminate across tones and group them according to these parameters. When only grouping by one parameter, participants did well. But when asked to group by all three parameters discrimination between tones suffered. When the experimenters told the participants to ignore one of the three acoustic parameters and pick just two parameters to rely on, participant performance increased for segregating the tones. However, the interesting finding in this study was that the two parameters chosen by participants to complete the grouping task seemed to be equally random pairings across participants. Kidd and Watson (see Watson & Kidd, 1994) suggest that there must be some subjective perceptual differences that participants rely on when interacting with acoustic parameters of auditory displays.

1.5.1 Musical Experience

One of the subjective qualities that has been explored among both stream segregation (Drake & Botte, 1993; Huron, 1989; Jones, Jagacinski, Yee, Floyd, & Klapp, 1995; McAdams et al., 1995; Schuett, 2010;) and auditory display researchers (Lacherez, Seah, & Sanderson, 2007; Sanderson, Wee, & Lacherez, 2006; Smith & Walker; 2005; Walker & Nees, 2005; Watson & Kidd, 1994) is musical experience. In tasks with multiple acoustic streams, participants with higher levels of music experiences have shown greater detection sensitivity to small changes in tempo (Drake & Botte, 1993; Jones et al, 1995) and alteration in frequency groupings (Huron, 1989). Brochard, Drake, Botte, and McAdams (1999) created a task where listeners had to monitor three auditory streams, designed to group as low, middle, and high frequency ranges. Participants were instructed to listen for small temporal irregularities within these complex streams and signal when they were detected. Overall, participants that were identified as musically trained (7+ years of experience and played an instrument at least 2 hours or more per week), performed better than non-musically trained listeners at detection accuracy. However, the researchers did not find any significant interaction between musical training and the other variables of frequency grouping or target location.

1.5.2 Training and Familiarity with Context

Lacherez et al. (2007) evaluated the learnability of melodic alarms used by anesthesiologists. These alarms represented high-priority events that often occur in a hospital setting such as alarms for changes in oxygen, ventilation, and cardiovascular. There were eight alarms in total, each represented with a melody. These alarms could overlap, so each melody was intended to act as its own auditory stream. Participants were active nurses with experience in anesthesiology, therefore it was hypothesized that their prior medical knowledge should help. This study was created as a follow-up to prior work exploring the same alarms with non-nurses (Sanderson et al, 2006).

The researchers trained participants on the alarms and the associated meanings, and the study took place over two days. On the first day participants were given a practice phase during which they could listen to the alarms up to 24 times while identifying the meanings. Following this practice was a test phase with feedback provided. On the second day, participants started with a retention check to see if they could identify the alarms meanings from memory. This was followed by a phase of four practice runs and a final alarm identification task. Listener accuracy increased with practice throughout the day, but did not improve across days.

Results differed between nurses that had one or more years of musical training compared to nurses with no musical background. Participants with musical training had higher accuracy overall, but did not show significantly faster response times to the stimuli. Lacherez et al. concluded that participants could not identify the alarm melodies with what the researchers considered to be enough overall reliability for the high-priority importance of the alarms in a hospital scenario, proposing that even with training a failure of stream segregation between the overlapping alarm melodies was limiting listener perception. The alarm melodies used in this study, they were all created in a musical scale of C minor. This was likely done for uniformity of the soundscape, but it means there was not much frequency separation between the tones in overlapping streams. Therefore, the unsatisfactory reliability in alarm identification is likely due to lack of frequency separation.

Comparing results from both Lacherez et al. (2007) who used nurses as their participants and Sanderson et al. (2006) who used non-nursing students, both nurses and non-nurses who reported at least one year of musical training had better identification of the alarm melodies than participants without musical training. While this shared finding supports the hypothesis that musical training aids in perception of multi-stream sonifications, there is a large difference between the one year of training found by Lacherez et al. (2007) and Sanderson et al. (2006) compared to the 7+ years of musical experience found in the results of Brochard et al. (1999). In addition to this variance, there are many other studies not talked about in this document that find no evidence for the impact of musical training. Walker and Nees (2011) suggest that there may be too much variation in what researchers and participants consider musical experience. Even musical ability may vary between participants who have both had the same amount of time practicing on the same type of instrument. Watson and Kidd (1994) suggest a slightly different explanation for observed correlations between musical experience and performance on these types of tasks. They propose that people who become musicians are a self-selecting population. For instance, someone who has very bad pitch perception is not as likely to become a musician, so the population of musically trained listeners may have already removed outliers with low perceptual ability.

1.5.3 Practice

A correlation between past musical experience and participants' ability to analytically listen to a complex group of sounds in an auditory display might indicate that listeners can learn to better segregate auditory streams. Although the link between musical experience and stream segregation is somewhat unclear, the overarching idea behind it implies that some form of familiarity with the acoustic properties useful to the listener will help with stream segregation. Thus, training with a complex or multi-stream sonification should help increase usability by teaching the listener how information streams are being sonified.

As discussed earlier in this document, for a sonification to be useful, context about the values being represented is important (Nees & Walker, 2007; Walker & Nees, 2011). Smith and Walker (2005) determined that for a sonification to provide some level of detail to the user, point estimation must be an option. That is to say, listeners may want to know what a specific value is, in order to compare it to another value in the sonification or to make some inferred decision (e.g., when I hear my stock reach $50 I will sell it). Smith and Walker evaluated listener performance on point estimation tasks to determine if listeners could more accurately identify point values in a sonification with the addition of x-axis and y-axis context as well as training.

The researchers found that both contextual features and training helped reduce listener error in point estimation, however these effects were not seen in an interaction. In other words, there was no additive effect on reducing listener error when both training and contextual features were implemented. Smith and Walker admit that while training was found to help listeners, it was unclear if this was due to the "sharpening [of] sensory skills" or simply the effect of practice when listeners received knowledge of results. Either way, training in this study helped listeners to better utilize the information imbedded within the target sonification stream.

To extend the findings of Smith and Walker (2005) and provide clarity on the role of training, Walker and Nees (2005) created five different experimental training conditions (control – no practice, practice only, practice with feedback, practice with visual prompts, and conceptual training). After participants completed their assigned training condition, they were given the same point estimation task used by Smith and Walker (2005). Results showed that practice with feedback lead to significantly lower error. Thus, having knowledge of results during training can have an impact on listener performance.

There is clearly benefit to having some form of acoustic experience similar in properties to that of the target stream. If, for example, the listener has years of experience playing the violin and the sonification they are interacting with is relying on timbre similar to that of a violin, then it makes sense that their perceptual and cognitive ability is pre-tuned for analytical listening in that frequency range and timbre. This is what Watson and Kidd (1994) call the musician's "trained ear." However, the benefit of familiarity or experience with those specific acoustic parameters is only useful if the listener knows that the target stream has those properties. With unfamiliar stimuli or sound sources, a trained listener will perform no better than a novice one. That said, training can impact listeners' ability to utilize complex sonifications if that training is specific to the auditory parameters that will be used in the auditory display and if that training includes feedback.

1.6 Type of Auditory Display Task

The way in which listeners are intended to interact with a sonification has an impact on how many acoustic variables listeners can utilize. The way information is displayed through an auditory display should be relevant and useful for the type of task the user will be performing (Barrass, 1997; Frysinger, 1990). This is similar to the way sounds being used to represent or replace images and icons in multimodal interfaces should fit with the role they are intended to play (Barrass, 2005; Brewster, 1994; Mezrich, Frysinger, & Slivjanovski, 1984). The types of tasks that listeners intend to perform using sonified data can be categorized in a few different ways. Walker and Nees (2011) discuss seven different task types: *monitoring, awareness of a process or situation, data exploration, point estimation or point comparison, trend identification*, and *exploratory inspection.* While these all have specific goals and subtasks for the listener, Kramer (1994b; Barrass & Kramer, 1999) describes all of these tasks as falling into two primary categories, monitoring tasks and analysis tasks. We will discuss tasks using these two higher order categorizations because it may be easier to consider the role of multiple streams and compare between past work under these headings rather than the more detailed categorization.

1.6.1 Monitoring Tasks

Monitoring includes a variety of tasks during which the listener attends a sonification over a period of time. This may include listening for a change to occur in the data, listening to status indications for awareness of a process or situation, or waiting until the auditory stream reaches a specific value. For these types of tasks, the display designer is relying on listeners' ability to detect small changes in auditory patterns or streams over time (Kramer et al., 2010). These types of sonifications have been designed for status monitoring in factory settings (Gaver, Smith, & O'Shea, 1991; Walker & Kramer, 1996; 2005) as well as for monitoring patients' vital signs in a hospital setting (Anderson & Sanderson 2004; 2009; Fitch & Kramer, 1994; Loeb & Fitch, 2002; Sanderson, Watson, & Russell, 2005). While much of the empirical work with multi-stream sonifications in anesthesiologist workstations has already been discussed in this document (see Sonification Configuration and Musical Experience) the premise for the task in these studies with regard to multiple auditory streams is relatively clear. Listeners are asked to monitor multiple auditory streams, differing on various acoustic parameters, and when a change in the pattern occurs the listener identifies it. Changes could occur on any of the acoustic parameters, such as an alteration in tempo or shift in frequency, and the listener can identify this acoustic change as an alteration in the situation they were tasked with monitoring. These multi-stream sonification studies provide evidence that listeners can successfully utilize up to eight different data variables (Fitch & Kramer, 1994) and up to three separate auditory streams (Anderson & Sanderson, 2004) in monitoring tasks of complex sonifications.

1.6.2 Analysis Tasks

Analysis tasks focus more on data exploration, the listener does not generally have a set template of what to expect, but must instead make sense of the relationship between the embedded sounds and the data values that are represented by the sonification. These types of tasks include data exploration (Brown & Brewster, 2003; Brown, Brewster, Ramloll, Yu, & Riedel, 2002; Flowers, Buhman, & Turnage, 1997; Flowers, Whitwer, Grafel, & Kotan, 2001; Hayward, 1994), point estimation or point comparisons (Bonebright, Nees, Connerley, & McCain, 2001; Smith & Walker, 2005; Walker, 2000; Walker & Lane, 2001; Walker & Nees, 2005), or the identification of trends in the data (Nees & Walker, 2008; Schuett et al., 2014; Song & Beilharz, 2007).

These different analytical tasks have slightly different goals for the listener, but the overall task of attending multiple streams while understanding the sounds and their values remain similar. In monitoring type tasks, the listener may not need to fully segregate between streams until they identify that a change has occurred. At that point the listener can direct their attention to the new target stream and utilize the information. With the analysis task, the listener may have some familiarity with the context of what they are listening to, but the relationship of acoustic elements within and between streams may be harder to segregate because they may not know how streams are supposed to form in relation to the information being provided. If this occurs we may see a decrease in listener performance when too many streams or variables are represented.

This may not be the case as we look for evidence to support the utility of multiple streams in analysis tasks. Sonifications of auditory graphs have resulted in listeners being able to successfully use at least two streams (Bonebright et al., 2001; Brown & Brewster, 2003; Song & Beilharz, 2007) with what are essentially two variables of the x-axis and y-axis values (Smith & Walker, 2002). A more recent study (Schuett et al., 2014) applied the sonification configuration technique, usually implemented for multi-stream monitoring tasks (Anderson & Sanderson, 2004), to a trend identification analysis task that included five different weather and climate variables. The researchers were able to map these five variables on up to three streams. Participants listened to the multi-stream sonification without knowing which of the five variables they would be asked about, and then reported on the trend in values for whichever of the variables was identified as the target during the questioning phase of each trial. Although participants had a set number of options (because the data was based on realistic weather patterns) they showed better than chance accuracy at identifying trends in weather variables when all five variables were presented concurrently on three streams within the sonification. While this may suggest that listeners in analysis tasks could utilize three auditory streams, all prior literature tends to show a limit of about two auditory streams for listeners with these types of sonifications. Further research is needed to investigate stream limitations for these types of sonification tasks.

Perhaps it is more likely that the number of streams listeners could utilize in the weather sonification study was more the result of proper display design and configuration rather than task based.

1.7 Summary

Summarizing what has been discussed in this document to answer the question — How many auditory streams can listeners utilize within a sonification? — we know that this answer is dependent on a few factors. Specifically, type of task, auditory display design choices, and training or individual differences of the listeners. Empirical work using auditory displays for monitoring tasks, such as keeping track of patients' vitals in a hospital (Anderson & Sanderson, 2004; 2009; Fitch & Kramer, 1994; Loeb & Fitch, 2002; Sanderson, Watson, & Russell, 2005) or listening for status alerts in a factory (Gaver, Smith, & O'Shea, 1991; Walker & Kramer, 1996; 2005), have identified the upper limit of eight data variables across three different auditory streams. Analysis tasks have not pushed this limit of auditory streams as high, with only one somewhat recent study (Schuett et al., 2014) intentionally measuring the utility of up to three auditory streams within a sonification.

Display design choices on how to map data variables to acoustic parameters and auditory streams, discussed earlier as sonification configuration, can help create a more ideal auditory display for a given context. However, this intentional design, often created by someone with musical background, is not always feasible. Sometimes a display is intended for general use or needs to be flexible in representing values across different domains. In other scenarios there may not be enough naturalistic sounds to represent a concept without applying arbitrary mapping between abstract sounds and real-world variables. Thus, it is important that we explore the utility of complex multi-stream sonifications in scenarios that also include less than ideal mappings.

Individual differences in musical or acoustic experience have had some measurable impact on listener accuracy for sonification tasks, but this has been hard to quantify between empirical studies since it is not usually measured in a standardized way. A better understanding of these differences, as well as how much they impact listener performance, is useful in two ways. First, knowledge about the factors that contribute to listener success with sonifications allows us to design better auditory displays based on who will be using them. Second, for high priority tasks that require auditory displays, knowing which users are be best suited to train on and use the displays could help save valuable time and resources in medical or military scenarios.

The study in this dissertation continues this program of research, building upon the results of Schuett et al. (2014) by investigating listeners' ability to utilize multi-stream sonifications in a trend identification task. Asking listeners to discern multiple data trends simultaneously is a somewhat difficult analytical task, but this is ideal because it tests listeners' performance. This study considers the effect of practice with feedback for aiding listener accuracy when identifying targets among multiple acoustic streams (Walker & Nees, 2005) and includes a practice phase between evaluation tasks. This allows us to investigate initial listener performance as well as performance after practice for this complex sonification analysis task. Additionally, the importance of display mapping choices was evaluated through four between-subjects display mapping conditions. And finally, the impact of listener differences, such as overall listening discernibility, musical experience, and motivation during the listening task and practice phase, was measured to determine how well these factors can predict overall listener performance on the initial display task and how well these factors can predict listeners' overall improvement through practice.

1.8 Research Questions

In this dissertation I seek to provide a deeper understanding of the effects that display design and individual listener differences have on listeners ability to utilize multi-stream or complex sonifications for analytical listening tasks. More specifically I set out to answer the following research questions:

1. Can the results of Schuett et al. (2014) be replicated to show that listeners can utilize five variables represented simultaneously in a sonification for trend identification?
2. How does listener accuracy change for multi-stream sonifications when the same acoustic parameters are mapped to represent different domains and context?
3. How well does a standardized listener experience score and study specific acoustic discrimination score predict participants' ability to utilize a multi-stream sonification?

In addition to these research questions I measured the impact of practice by recording participant accuracy in the trend identification task at two time-points. The first test evaluates how initially easy the sonification is to use, the second test evaluates listener accuracy on the trend identification task after having time to practice with the display mapping. Participants were asked to rate their motivation during the evaluation and practice phases of the study. Because the sonification task was designed to push the upper limits of stream segregation, participants were presented with a challenging task. Measuring motivation throughout the study acts as an additional control.

CHAPTER 2. EXPERIMENT

An empirical study was conducted to answer the research questions proposed in the previous section. This investigation evaluates listeners' ability to utilize multi-stream sonifications for trend identification. This approach creates an analytical listening task that requires listeners to use a sonification containing five different acoustic parameters. We know from findings discussed in Section 1.1.5 that listeners struggle to utilize five individual acoustic parameters simultaneously, but they will likely group these parameters into fewer auditory streams. Considerations based on sonification configuration (Anderson & Sanderson, 2004) were used in the auditory display design for this study, composing acoustic parameters to encourage ideal stream segregation.

The impact of display mapping design choices on listener accuracy was evaluated using the same acoustic parameters for representation of two different context domains. These context domains were weather variables (temperature, humidity, wind speed, air pressure, and cloud cover) and health variables (blood oxygen level, respiratory rate, heart rate, blood pressure, and body temperature). Conditions within these context domains were designed to measure the effect of intentional mapping, where display experts were asked to assign domain variables to each acoustic parameter based on rationale, versus an arbitrary assignment of each domain variable to one of the acoustic parameters. The acoustic parameters used in these sonifications were frequency, tremolo, intensity, pink noise intensity, and a filtered noise. These parameters were originally selected as best fits for the weather domain.

Participants’ ability to utilize these display conditions was measured with accuracy scores during two evaluation tasks. The initial evaluation occurred at the beginning of the study to assess participants’ initial ability to utilize the sonification for trend identification among the five domain specific variables they were assigned. As discussed in Section 1.6.3 it is beneficial for listeners to practice and receive feedback to reduce error in analytical sonification tasks (Walker & Nees, 2005), so a second evaluation occurred after participants were given time to practice with the sonifications. Using a pre- and post-test approach for listener evaluations allows the results of this study to include a measure of initial listener accuracy when first using a complex sonification, as well as a measure of listener accuracy when using the complex sonification with familiarity.

2.1 Method

2.1.1 Participants

Participants for this study included 103 members of the Georgia Tech community, ranging in age from 18 to 28 years old (mean age = 19.76, SD = 1.72). Participants were recruited through the Georgia Tech SONA system and through word of mouth and self-reported normal or corrected-to-normal hearing and vision. Three participants are not included in the analysis; one participant asked to leave before completion and the other two did not follow instructions correctly.

2.1.2 Materials

2.1.2.1 <u>Listening Discrimination Task</u>

As discussed in Section 1.5, one of the often-cited reasons behind differences in individual performance is that some participants have a “trained ear.” In other words, some participants have an enhanced ability to discern the differences between acoustic stimuli. Whether this is learned through musical training, prolonged exposure to music, or is an innate gift, this listener difference can be accounted for if we measure it. Because the primary task in this study required listeners to utilize multiple concurrent auditory streams, the listening discrimination task consists of multiple concurrent audio elements.

The Listening Discrimination Task required participants to listen to one audio track, followed by another, and determine if the first and second track are the same or different. The first and second track were either the same, or differ by one acoustic parameter each time. The task increased in difficulty as the number of acoustic parameters in the track increased. When there is only one acoustic parameter and then the second track presents a change across that single parameter it is relatively easy for the listener to discern the difference. But when there are multiple acoustic parameters in each track, detecting the presence of a change becomes increasingly more difficult.

For each trial, participants were presented Track A followed by Track B and given time to respond by saying “same” or “different.” This task consisted of 26 total trials, in half of the trials tracks A and B are the same, and for the other half they are different. The trial difficulty was presented in a randomized order for each participant. The acoustic parameters used for each of the thirteen acoustic groupings are included in Appendix A.

2.1.2.2 Musical Sophistication Index

Musical experience is often debated as a factor that can account for performance differences between listeners. While this connection has never been fully supported (Walker & Nees, 2001; Watson & Kidd, 1994) the current study includes a metric of musical sophistication for participants using Goldsmiths Musical Sophistication Index (Gold-MSI), asking participants to self-report musical skills and behaviors (Müllensiefen, Gingras, Musil, & Stewart, 2014a; 2014b). This assessment was chosen because it includes more than listeners' history with musical instruments. It includes a variety of items that measure listeners' overall level of musical sophistication. This metric is based on questions about musical engagement, self-reported perceptual abilities, amount of past musical training, and emotional connection with music. A copy of the questionnaire items is provided in Appendix B. This measure was used to create a musical sophistication score for each participant, discussed in the results section.

2.1.2.3 Intrinsic Motivation Inventory

Throughout the study participants were asked to complete the Intrinsic Motivation Inventory (IMI) scale (Deci, Eghrari, Patrick, & Leone, 1994; Plant & Ryan, 1985; Ryan, 1982; Ryan, Mims & Koestner, 1983). This scale is used as a measure for subjective motivation towards a specific task during laboratory studies. The first motivation check occurred immediately after the first evaluation. This was included to gauge listeners' motivation during the evaluation phase. The second motivation check occurred at the end of training. At this point participants were reporting their level of motivation on the task after having completed the somewhat lengthy practice phase. The third motivation check occurred immediately after the final evaluation.

The purpose of these motivation checks is to determine if the participants got bored, tired, or distracted at any point during one of the evaluation or training phases. It relies on self-reported participant responses, but nevertheless offers some level of measure. The Intrinsic Motivation Inventory is included in Appendix C. The motivation scores for each participant initially act as a check to potentially weed out any participants who did not take the study seriously or lacked motivation to try during the practice phase. These scores also offer the opportunity to look for a correlation between participants' motivation scores and their score increases between evaluations 1 and 2.

2.1.3 Display Design & Mapping

The four display-mappings used in this study were predicted to range from best to worse based on a combination of the data domain and the design mapping. The sonifications used in these conditions were created to facilitate the presentation of five separate domain variables across three distinct auditory streams. All four mappings use the same acoustic variables and each condition group utilized tracks from the same prerecorded playlist. The auditory streams were separated spatially by panning parameters into the left ear, the right ear, and a centered acoustic stream present in both ears. The centered stream is used to only represent one variable, while the left- and right-panned streams include two parameters or variables on each. This decision was made to optimize the presentation of all five variables across three intended streams and encourage stream segregation for listeners. Table 1 summarizes the acoustic mapping of the sonifications. The left-panned stream includes changes in frequency of a sine wave as well as a tremolo or wavering effect caused by changes to the repetition rate of the sine wave. The center stream includes intensity changes on a chord. And the right-panned stream is made from intensity changes to pink noise, presented with a filter that can increase or decrease the spectrum.

Table 1. Acoustic Parameters Mapped to Panning

Left Pan	Centered	Right Pan
Frequency	Chord (Intensity changes)	Pink Noise (Intensity changes)
Tremolo		Filter (filter on pink noise)

The data trends represented by each of these five parameters can increase, decrease, remain constant, increase-then-decrease, or decrease-then-increase over time. The display is intended to represent informative trends that any of the weather or health parameters could have in a given snapshot of time. For instance, *humidity* or *heart rate* are both variables that can rise, fall, or remain constant over a period of time in the real world. The task remained the same for all conditions, differing only by which concepts participants are instructed to map to each acoustic parameter.

2.1.3.1 Expert Designed Weather Domain Mapping

The sonification configuration mapping that was hypothesized to be most ideal, was based on expert design with guidance from the results of a sound mapping study conducted as a lead-up to the current study display design. This study included 133 undergraduate students at the Georgia Institute of Technology, all with self-reported normal or corrected-to-normal hearing. Participants were seated in front of a computer and given a pair of SONY MDR-V150 Headphones. These listeners were asked to follow the instructions presented on a website that was used to present the sound and concept options.

The website consisted of a series of pages, in a randomized order, that asked participants to listen to each acoustic parameter and match it to the weather concept that it best represents. Results of this study are included in Appendix D.

The weather concepts consisted of *temperature, humidity, cloud cover, wind speed*, and *air pressure*. These specific variables were chosen because they all naturally follow similar data trends and have a natural relationship to how weather is physically experienced. Weather data is a context regularly used in sonification studies due to listeners' usual familiarity with the concepts and the variety of mappings (Flowers et al., 2001; Herman, Drees, & Ritter, 2003; Polli, 2004; Schuett et al., 2014).

The display design choices took into consideration all the results of the weather mapping study. In some cases, there was a clear mapping favorite among listeners, such as representing temperature changes with frequency or wind speed with intensity. For other weather concepts there was a less distinct winner between multiple acoustic options. These mapping results were shared with two expert sound designers and they were asked to create the current display. Sonification configuration for this display relies on representing multiple data concepts on the same auditory stream, and this seems to work better when the variables being represented together have a conceptual relationship (Schuett et al., 2014). For instance, temperature and humidity can affect the way each other is perceived. The two concepts have a physical relationship and thus make more sense to be mapped together on the same stream. Similarly, wind speed and air pressure are physically related because changes in air pressure can lead to increases or decreases in wind speed. For this reason, it made sense to represent these two concepts together on the same auditory stream. This left cloud cover as the final variable to be represented by itself because it has less of a physical connection with any of the other four variables. A summary of the ideal display mapping is shown in Table 2.

Table 2. Expert Designed Mapping with Weather Variables

Concept Variable	Acoustic Parameter	Stream
Temperature	Frequency	Left
Humidity	Tremolo	Left
Cloud Cover	Intensity (chord)	Center
Wind Speed	Pink Noise Intensity	Right
Air Pressure	Filter (on noise)	Right

2.1.3.2 Expert Designed Health Domain Mapping

The second concept-to-acoustic mapping is based on best fit and expert opinion. Unlike the weather display mapping, this display design is not based on the results of a sound-mapping listener study. Instead, the same acoustic parameters from the first display are used to represent five variables in a different context. The rationale for including this mapping condition is to represent a sonification with informed design, but not participatory data to support the mapping decisions.

The context of health data was chosen for this condition, congruent with past sonification configuration studies (Anderson & Sanderson, 2004; 2009; Fitch & Kramer, 1994; Lacherez et al., 2007; Loeb & Fitch, 2002; Sanderson, Watson, & Russell, 2005; Sanderson, Wee, & Lacherez, 2006). The specific variables selected for this study were *respiratory rate*, *heart rate*, *body temperature*, *blood oxygen level*, and *blood pressure*.

Because both the weather and health mapping conditions utilize the same display design, with different concept-to-acoustic mappings, the health domain mapping also required the pairing of two concepts for the left acoustic stream and two concepts for the right acoustic stream. The fifth concept was, again, represented by itself in the middle acoustic stream. Respiratory rate and heart rate were paired together because of the physical connection between the two concepts. Blood oxygen level and blood pressure were also paired due to real-world connection between these variables. Body temperature has the least direct physical connection with the real-world data fluctuation in the other four variables, since it acts as a more global indication of change in the body overall. Thus, body temperature was assigned its own auditory stream like cloud cover from the first mapping condition. Using the expert opinions of the sound designers, mappings for each of the health variables were applied to the established acoustic parameters. These mappings are listed in Table 3.

Table 3. Expert Designed Mapping with Health Variables.

Concept Variable	Acoustic Parameter	Stream
Respiratory Rate	Frequency	Left
Heart Rate	Tremolo	Left
Body Temperature	Intensity (chord)	Center
Blood Oxygen Level	Pink Noise Intensity	Right
Blood Pressure	Filter (on noise)	Right

2.1.3.3 Arbitrary Mapping of Weather and Health Variables

The first two display conditions represent an ideal design, based on listener feedback and sonification configuration, and a best-fit design, based on expert opinion and sonification configuration principles. To evaluate the effectiveness of expert opinions on display mapping and sonification configuration, the study included arbitrarily mapped conditions for both weather and health variables. The arbitrarily mapped weather display represents a mapping of variables to acoustic parameters that were previously rated by listeners as a good fit for the weather domain, but there was no display design consideration for how the weather variables are mapped to each of the acoustic parameters. The fourth display mapping with arbitrarily assigned health variables represents a control condition where the acoustic parameters were not at all selected with the health domain in mind and variables are somewhat randomly assigned to the display mapping. This would be the equivalent of a 'one-size-fits-all' approach to sonification design when representing data trends.

To allow for overall comparison between the weather and health contexts the arbitrarily mapped weather and health conditions keep the same weather-to-health relationship. In other words, the same mapping is applied to both conditions. The acoustic parameters were numbered 1 through 5 and a randomly generated pattern was used to map the concepts to this new arbitrary acoustic parameter order. Table 4 and Table 5 summarize the arbitrary weather and health display mappings.

Table 4. Arbitrary Display Mapping for Weather Variables.

Concept Variable	Acoustic Parameter	Stream
Wind Speed	Frequency	Left
Temperature	Tremolo	Left
Humidity	Intensity (chord)	Center
Air Pressure	Pink Noise Intensity	Right
Cloud Cover	Filter (on noise)	Right

Table 5. Arbitrary Display Mapping for Health Variables.

Concept Variable	Acoustic Parameter	Stream
Blood Oxygen Level	Frequency	Left
Respiratory Rate	Tremolo	Left
Heart Rate	Intensity (chord)	Center
Blood Pressure	Pink Noise Intensity	Right
Body Temperature	Filter (on noise)	Right

2.1.4 Procedure

Upon arrival participants were randomly assigned to one of the four display mapping conditions. The study uses a between-subjects design because once participants were familiarized with one of the display mappings it could lead to unintended confusion if they were asked to then learn another. Instead participants were assigned a mapping condition and all of their interactions with the auditory display are framed in terms of that condition's concept-to-acoustic mapping context. Before participants were familiarized with the sonification they were asked to complete the Listening Discrimination Task.

2.1.4.1 Listening Discrimination Task

As described earlier, participants were presented Track A and Track B then given time to respond by saying "same" or "different." This task consisted of 26 total trials, in half of the trials tracks A and B are the same, and for the other half they were different. This task took roughly ten minutes to complete.

2.1.4.2 Introduction to the Display Mapping

Once each participant completed the Listening Discrimination Task they were given an interactive introduction program. The program operates like a webpage, allowing a user to click through each of the concept-to-acoustic parameter mappings. The introduction program was specific to the display mapping condition groups assigned at the beginning of the study. Participants were able to listen to these mapping examples as many times as they liked, until feeling comfortable with the instructions. The display mapping introduction did not allow participants an opportunity to practice with the multi-stream version of the display; they simply got a chance to learn the mappings for each variable.

2.1.4.3 Evaluation 1

After the participant was comfortable with their introduction to the display they were given an evaluation task. This first test acts as an evaluation of the listeners' ability to comprehend the data representations within the display without any practice. When the participant indicated that they were ready to start this task the trials began. Participants used SONY MDR-V150 headphones and sat at a table located in a sound attenuated booth. Participants were given an answer sheet to record their responses. A copy of this answer sheet is included in Appendix E.

For each trial the participant listened to a ten second clip of the auditory display. Participants were instructed to listen to the data patterns for each of the five weather or health variables. For each trial all five variables are represented simultaneously within the display recording. When the audio clip finished, the experimenter asks each participant about one of the five variables. The participant used their answer sheet to circle the data pattern that was represented for that variable. For instance, if the participant was in the weather condition and the frequency was increasing during the audio clip in that trial, if the researcher asked the participant to report about ‘temperature’ the participant should correctly select the arrow indicating an increase over time. After each trial the participant would indicate that they are ready to advance to the next, they would again listen to a full clip and are asked to select the correct trend pattern for a different variable.

The evaluation test consisted of twenty trials, presented in a randomized order for each participant. For each trial the researcher asked the participant to recall one of the variables, the target variable for each trial was read out from one of twenty different pre-randomized lists. The target variable lists were randomized to ask about each of the five variables an equal number of times across all twenty trials. These randomized target variable orders were numbered one through five and updated to reflect the five variable orders used in each of the weather and health conditions.

2.1.4.4 Motivation Check 1

As soon as participants finish the first evaluation phase they were promptly asked to complete the Intrinsic Motivation Inventory (IMI) scale. They were also instructed to take a brief rest to avoid fatigue.

2.1.4.5 Practice with Feedback

Following a short rest break, participants began the practice phase. The initial practice task was very similar to the evaluation task. Participants listened to a display recording, similar to what they previously heard, but pulled from a different recording set. None of the practice audio clips were the same as the evaluation trial stimuli. Participants were asked about one of the variables, similar to the evaluation task, and provided an answer. The practice phase differs from the evaluation tests because the researcher provided participants with knowledge of their results. By providing feedback the listener would know if they were responding to answers correctly. If the participant was correct they would move on to another practice trial, if they were incorrect they were allowed to listen to the display clip again and re-answer.

On the second time listening to a practice trial, the participants already knew the target variable. This allowed them an opportunity to listen for specific acoustic parameters that they had difficulty with. This process was repeated until the listener got five correct responses in a row. Participants were strongly encouraged to indicate when they wanted to take a break, and there was a mandatory break in the middle of the practice phase to avoid listener fatigue. If the participant was able to successfully answer five practice trials in a row they were given an option to begin the second evaluation test or continue with practice. The practice phase lasted an average of 45 minutes, depending on the number of breaks and rest each listener wanted.

2.1.4.6 Motivation Check 2

Participants were asked to fill out the IMI scale at the end of their practice phase as a motivation check for the practice task. This also acted as a short break before the evaluation task started.

2.1.4.7 Evaluation 2

The second evaluation phase occurred after participants completed the practice trials. This evaluation followed the same procedure as the first evaluation, but with a randomized order of trials and target variables.

2.1.4.8 Motivation Check 3

Participants were asked to complete the third IMI scale as a motivation check immediately after the second evaluation trials.

2.1.4.9 Musical Sophistication Index

After each participant finished with the second evaluation motivation check they were asked to fill out a variant of the Goldsmiths Musical Sophistication Index.

2.2 Hypotheses

2.2.1 Hypothesis 1

Listeners will be able to successfully utilize a complex sonification of five variables for a trend identification task.

2.2.2 Hypothesis 2a

Participants assigned an intentionally (expert) designed sonification mapping will have higher performance scores compared to participants who are assigned to the arbitrarily mapped sonification conditions.

2.2.3 Hypothesis 2b

Participants assigned the Weather domain will have higher performance scores compared to participants who are assigned the Health domain. This is because the acoustic parameters were selected originally based on parody with weather variables.

2.2.4 Hypothesis 2c

There will be an interaction between Design and Domain. Participants who are in the intentionally designed weather mapping group will have highest performance scores, followed by those in the intentionally mapped health condition. Participants in the arbitrary display mapping groups will have the lowest scores, with the arbitrary health mapping display group having the lowest performance scores.

2.2.5 Hypothesis 3

Participants performance scores will increase with practice.

2.2.6 Hypothesis 4

Participants with higher scores on Musical Experience Index and Listening Discrimination Task will also have higher performance overall on the multi-stream sonification trend identification tasks.

CHAPTER 3. RESULTS

Participants' ability to utilize the complex five-variable sonifications within a trend identification task were measured by creating scores for the Evaluation 1 and Evaluation 2 tasks. Evaluation scores were calculated by adding the number of correctly identified trends across all twenty trials. Descriptive statistics for Evaluation 1 and Evaluation 2 scores are shown in Table 6. Results were analyzed using a 2 (display design) x 2 (display domain) mixed-design analysis of variance (ANOVA).

Table 6. Descriptive Statistics for Evaluation 1 and Evaluation 2 Scores.

	Mean	Std. Deviation	Min	Max	N
Eval 1 Scores	8.90	2.65	4	15	100
Eval 2 Scores	11.13	2.52	5	16	100

There is a significant main effect for display domain, $F(1, 96) = 3.97$, $p < .05$, ($\eta_p^2 = .04$), with higher mean accuracy scores for listeners in the weather conditions ($M = 10.42$) compared to mean accuracy scores for listeners in the health sonification conditions ($M = 9.61$). Figure 8 shows mean evaluation scores across display domain conditions. There was also a significant main effect for display design, $F(1, 96) = 8.56$, $p < .05$, ($\eta_p^2 = .082$), with higher mean listener accuracy scores in the expert designed conditions ($M = 10.61$) than those in the arbitrarily mapped conditions ($M = 9.42$). There was no statistically significant interaction between domain and design conditions $F(1, 96) = .001$, $p = .98$. Figure 9 shows mean evaluation scores across display design conditions.

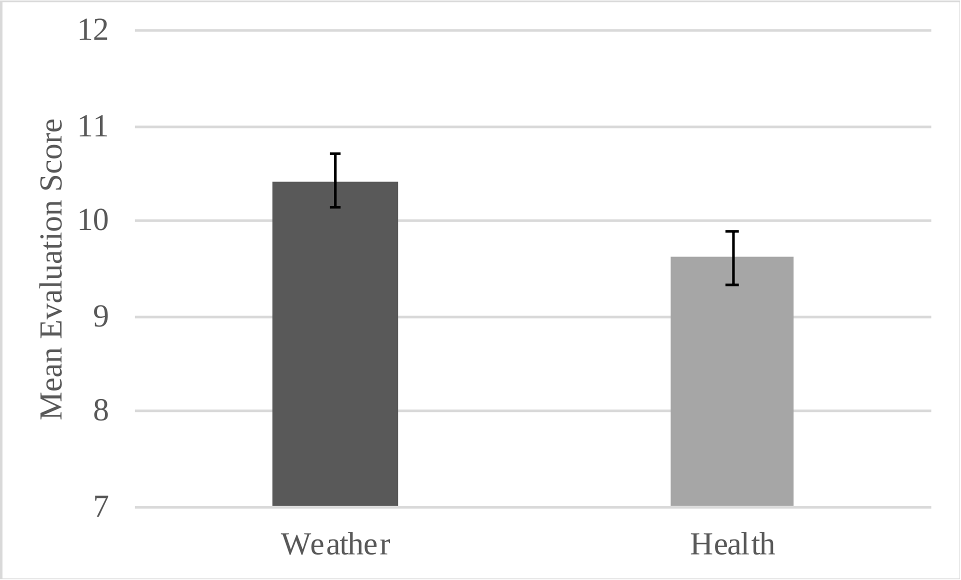


Figure 8. Mean evaluation scores between display domain conditions. Error bars indicate standard error of the mean.

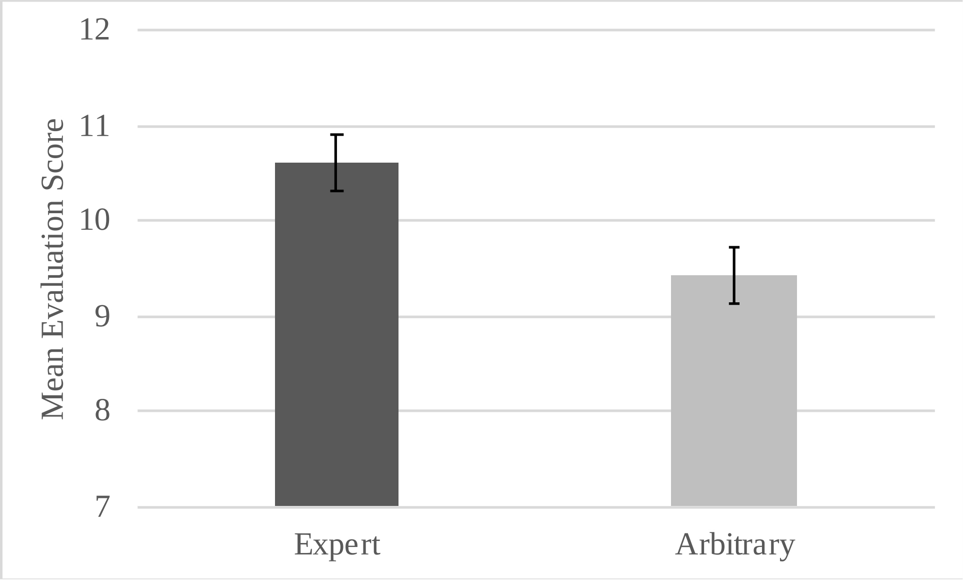


Figure 9. Mean evaluation scores between display design conditions. Error bars indicate standard error of the mean.

There is a main effect of practice, F(1, 96) = 57.535, $p < .05$, ($\eta_p^2 = .375$), with significant increase in mean accuracy scores from Evaluation 1 (Mean = 8.9, SD = 2.653) to Evaluation 2 (Mean = 11.13, SD = 2.521). Figure 10 shows the mean scores for both evaluation time-points. There was not a statistically significant interaction with practice for domain F(1, 96) = 1.417, p = .24, nor design conditions F(1, 96) = .97, p = .326. To visualize the effect of practice across each design and domain mapping condition group, mean listener accuracy scores are plotted in Figure 11. A post hoc test using a Bonferroni correction revealed a statistically significant difference between mean listener scores for the Expert Designed Weather Sonification condition (M= 11.02) and the Arbitrarily Assigned Health Mapping condition (M= 9.02), $p < .0125$. This was the only statistically significant difference as an effect of the display domain and mapping design.

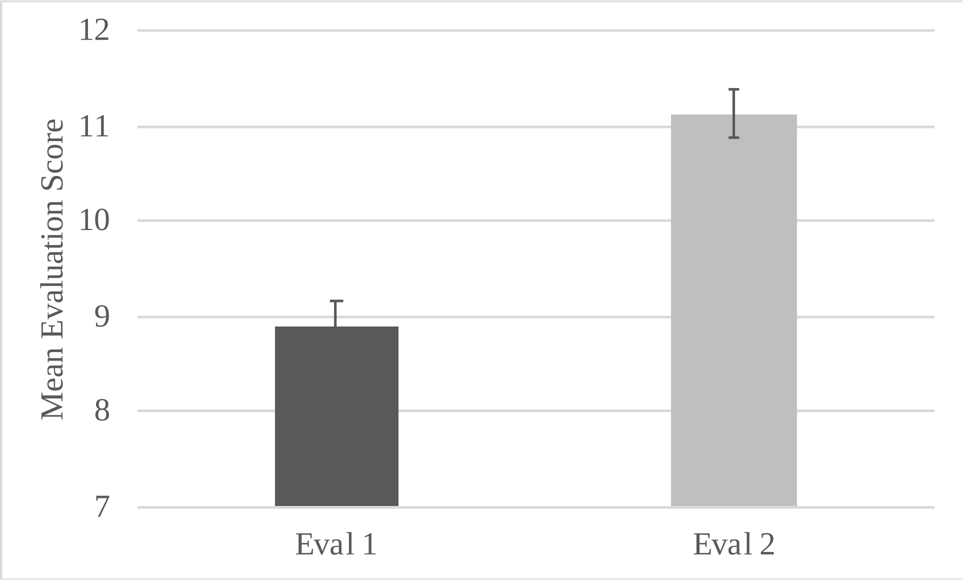


Figure 10. Mean scores between Evaluation 1 and Evaluation 2 as a result of practice. Error bars indicate standard error of the mean.

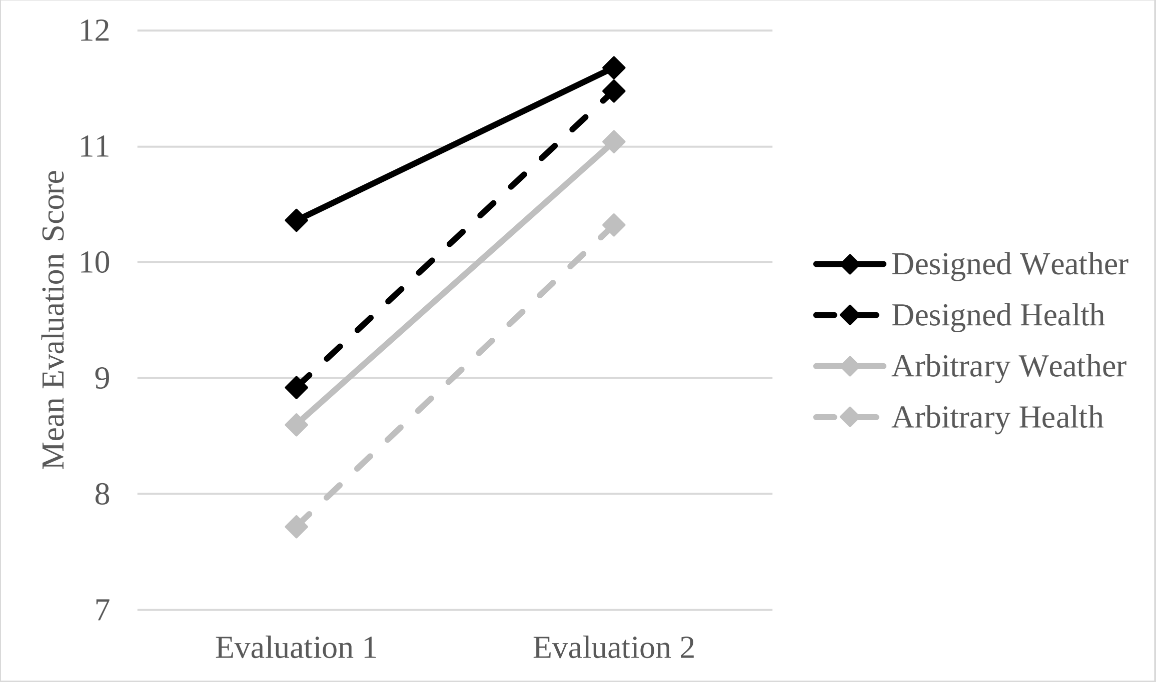


Figure 11. Mean evaluation scores plotted for each combination of Design and Domain conditions.

3.1 Intrinsic Motivation Inventory (IMI)

The Intrinsic Motivation Inventory (IMI), found in Appendix C, is made up of 22 items. Each item is a statement that participants were asked to rate on a 7-point scale, ranging from 1 (not at all true) to 7 (very true). Items 6, 11, 13, 14, and 18 are negatively worded and were reverse scored before analyzing results. A total IMI score was created by adding up scores across all 22 items for each participant at each of the motivation check time-points. The lowest possible motivation score on this scale is 22, the highest possible motivation score is 154.

A repeated measure analysis of variance (ANOVA) was conducted to compare mean IMI scores across all three motivation checks. Mean IMI scores differed between motivation check time-points with statistical significance, F(2, 196) = 66.79, $p < .05$, ($\eta_p^2 =$

.41). Pairwise comparisons between motivation check time-points was conducted using a Bonferroni adjustment. The mean IMI score from the second motivation check (M = 105.53, SD = 16.03) is statistically greater than both the mean IMI score for the first motivation check (M = 93.115, SD = 15.74) and mean IMI score for the third motivation check (M = 97.56, SD = 18.07), p< .0167. Mean IMI score from the third motivation check (M = 97.56, SD = 18.07) is great that mean IMI scores from the first motivation check (M = 93.115, SD = 15.74) with statistical significance, p< .0167. Mean IMI scores across all three motivation check time-points are plotted in Figure 12.

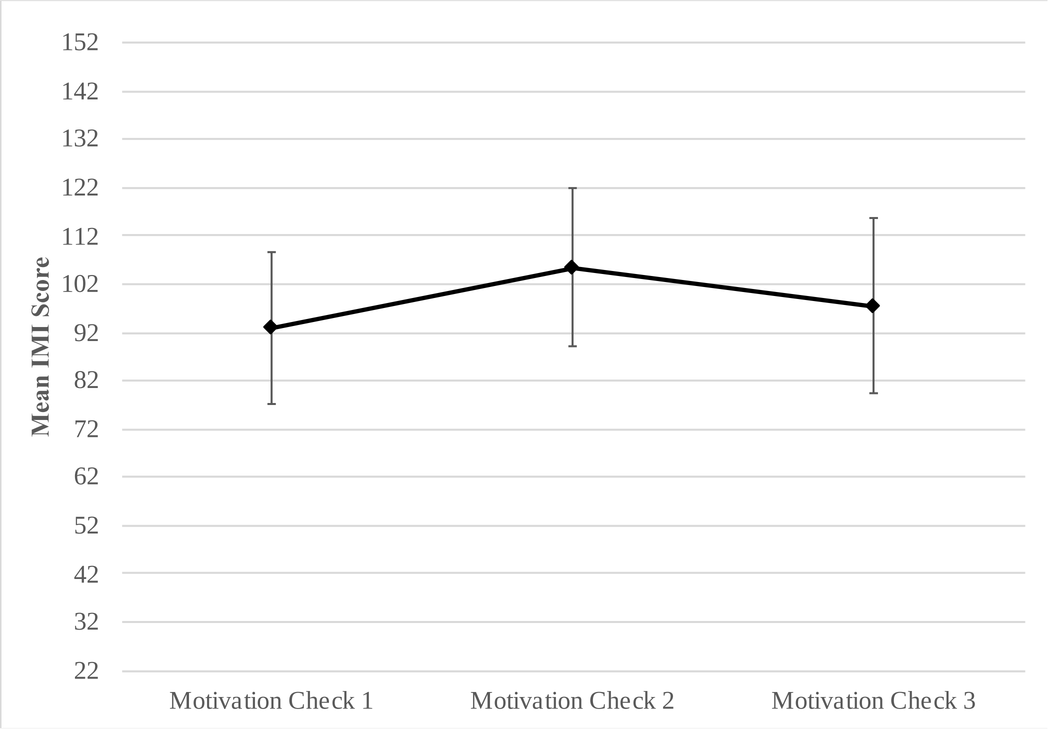


Figure 12. Mean Intrinsic Motivation Inventory (IMI) scores and standard deviations across the three motivation check time-points.

3.2 Listening Discrimination Task Score

Each participant began their session with a Listening Discrimination Test. There was a total of 26 trials that varied in difficulty. Listening Discrimination Scores were calculated for each participant by totaling the number of correct responses. Listening Discrimination Score across all participants had a Mean = 19.81, SD = 2.1, with individual responses ranging from 14 to 24.

3.3 Musical Sophistication Index (MSI) Score

The Musical Sophistication Index (MSI) consists of 15 self-report rating items on a 7-point Likert scale plus 3 open ended items. The MSI scale is included in Appendix B. Items 7, 9, 11, 13, and 14 are negatively worded and had to be reverse scored before adding up participant responses into an overall score. Musical Sophistication Index score across all participants had a Mean = 71.99, SD = 22.71, with individual responses ranging from 17 to 121.

3.4 Predicting Listener Performance on Trend Identification Task

A multiple linear regression was calculated to predict listener performance scores during the initial evaluation (Eval 1) based on Musical Sophistication Index (MSI) score, First Motivation Score (IMI1), and Listening Discrimination Task score (LD). A significant regression equation was found ($F(3, 96) = 4.267$, $p < .05$), with an R^2 of .118. Participants' predicted Eval 1 score is equal to 3.037 + .026 (MSI) + .037 (IMI1) + .025 (LD). As Musical Sophistication Index score increases by one standard deviation (23.31 points), Evaluation 1 score increases by .231 standard deviations. This interpretation is true if the effects of IMI1 and Listener Discrimination score are held constant. As listener IMI scores for the first motivation check increase by one standard deviation (15.485 points), Evaluation 1 score increases by .218 standard deviations if the effects of MSI score and Listener Discrimination score are held constant. As Listener Discrimination score increases by one standard deviation (2.097 points), Evaluation 1 score increases by .02 standard deviations when the effects of MSI score and IMI1 scores are held constant. Both Musical Sophistication Index score and IMI score for motivation during the first evaluation are significant predictors of Evaluation 1 scores. Table 7 includes a summary of the analysis variables. Table 8 is included to provide an overview of the descriptive statistics for these variables.

Table 7. Summary of Regression Analysis for Variables Predicting Evaluation 1 Scores (N=100)

	B	*SE B*	β	R Square Change
Constant	3.04	2.80		
Musical Sophistication Index Score	0.03	0.01	0.23*	.06*
Motivation 1 (IMI1)	0.04	0.02	0.22*	.04*
Listening Discrimination Task	0.03	0.12	0.02	.00

Notes. R^2 = .12, p< .05. *p< .05

Table 8. Descriptive Statistics for Variables Predicting Evaluation 1 Scores

	Mean	Std. Deviation	N
Evaluation 1 Score	8.90	2.65	100
Musical Sophistication Index Score	72.11	23.31	100
Motivation 1 (IMI1)	93.07	15.49	100
Listening Discrimination Task	19.81	2.10	100

A multiple linear regression was calculated to predict listener performance scores during the final evaluation (Eval 2) based on Musical Sophistication Index (MSI) score, participant motivation score after the practice phase (IMI2), motivation score after the final evaluation (IMI3), and Listener Discrimination Task score (LD). A significant regression equation was found (F(4, 94) = 7.518, p< .05), with an R^2 of .242. Participants' predicted Eval 2 score is equal to -.661 + .03 (MSI) - 0.16 (IMI2) + .027 (IMI3) + .436 (LD). As Musical Sophistication Index score increases by one standard deviation (23.4087 points), Evaluation 2 score increases by .276 standard deviations if the effects of IMI2 score, IMI3 score, and Listener Discrimination score are held constant. As listener IMI2 score increases by one standard deviation (16.027 points), Evaluation 2 score will decrease by -.101 standard deviations when the effects of MSI score, IMI3 score, and Listener Discrimination score are held constant. As listener IMI3 score increases by one standard deviation (18.066 points), Evaluation 2 score increases by .195 standard deviation if the effects of MSI score, IMI2, and Listener Discrimination score are held constant. As Listener Discrimination score increases by one standard deviation (2.072 points), Evaluation 2 score increases by .358 standard deviations if the effects of MSI score, IMI2, and IMI3 are held constant. Both Musical Sophistication Index score and Listener Discrimination score are significant predictors of Evaluation 2 scores. Motivation scores during practice (IMI2) and during the final evaluation task (IMI3) are not statistically significant predictors of Evaluation 2 score. Table 9 includes a summary of the analysis variables. Table 10 is included to provide an overview of the descriptive statistics for these variables.

Table 9. Summary of Regression Analysis for Variables Predicting Evaluation 2 Scores (N=99)

	B	*SE B*	β	R Square Change
Constant	-0.62	2.49		
Musical Sophistication Index Score	0.03	0.01	0.28*	.06*
Motivation 2 (IMI2)	-0.02	0.03	-0.10	.01
Motivation 3 (IMI3)	0.03	0.03	0.20	.00
Listening Discrimination Task	0.44	0.11	0.36*	.13*

Notes. $R^2 = .24$, p< .05. *p< .05

Table 10. Descriptive Statistics for Variables Predicting Evaluation 2 Scores

	Mean	Std. Deviation	N
Evaluation 2 Score	11.15	2.53	99
Musical Sophistication Index Score	72.02	23.41	99
Motivation 2 (IMI2)	105.52	16.03	99
Motivation 3 (IMI3)	97.51	18.07	99
Listening Discrimination Task	19.85	2.07	99

Note. A participant did not fill out the IMI3 and was excluded from analysis.

CHAPTER 4. DISCUSSION

4.1 Utilization of Complex Sonifications for Trend Identification Tasks

Overall participant scores on the evaluation provide evidence that listeners are able to utilize a multi-stream sonification well enough to correctly identify trends across five simultaneously displayed variables. Without any practice, listeners had a 44.5% accuracy while using the display. With less than an hour of practice, listeners could utilize a five-variable sonification in a trend identification task with 55.7% accuracy. Participants did not know what variable they would be asked about until the sonification ended, so the task required listeners to comprehend as many variables as possible during each trial. This evidence shows that listeners can utilize up to 5 acoustic parameters within a single sonification to complete an analytical listening task.

Although these results show that listeners can indeed utilize a multi-stream sonification, an overall mean accuracy of 55.7% may not indicate a practical success. However, we should consider the deliberate complexity of the listening task. In order to measure listeners' comprehension of multiple variables within the display, listeners were only allowed to listen to each display once during the evaluation tasks. Additionally, in each trial the sonification plays for a duration of 10 seconds, the equivalent task for a visual set of graphs would allow the user to study a graph of five variables for 10 seconds before removing the image and then asking about a trend in one of the variables. This challenge was discussed by Schuett & Walker (2013) when comparing listening comprehension tasks to information solicitation techniques when studying situation awareness.

Lacherez et al. (2007) did not consider their finding of 64% accuracy a reliable enough finding for practical implementation in a healthcare setting when asking musically trained nurses to correctly identify alarms in a monitoring task. However, the trend identification task used in this dissertation is an analytical listening task and requires listeners to identify not only which variables are changing but to also understand the data pattern represented by each change. Schuett et al. (2014) used a very similar trend identification task for a sonification used to represent only weather variables. Their results found a 64.22% accuracy, which is currently the highest accuracy for tasks in an empirical study that use a five-variable sonification. The listener accuracy reported by Schuett et al. (2014) may have been higher than what was found among participants in this dissertation because they used spatialized audio within their sonification design, but could also be a result of repeating stimuli across multiple trials while randomizing target variables.

If we are to use the 64% accuracy findings from Schuett et al. (2014) as a comparison of success for the study in this dissertation, 10% of participants had 65% or higher accuracy on the first evaluation, and 30% of participants had 65% or higher accuracy during the second evaluation. Figure 13 and Figure 14 show the distribution of evaluation score, by percentage correct, for Evaluation 1 and Evaluation 2 respectively. We see that after having time to practice with the sonification, by Evaluation 2 the top 10% of participants have accuracy of at least 75% correct. While it is difficult to say if this accuracy is acceptable in a practical setting, because that would depend on the real-world implementation, these results suggest that additional practice may continue to increase listener accuracy for this type of complex listening task.

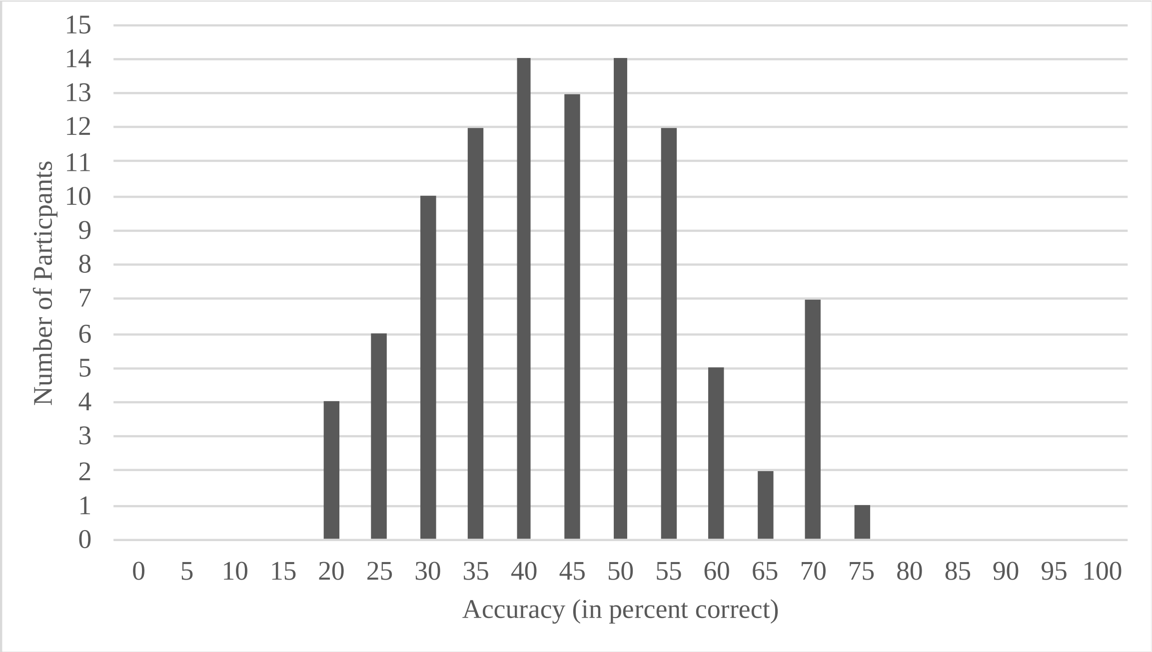


Figure 13. Distribution of participant accuracy for Evaluation 1.

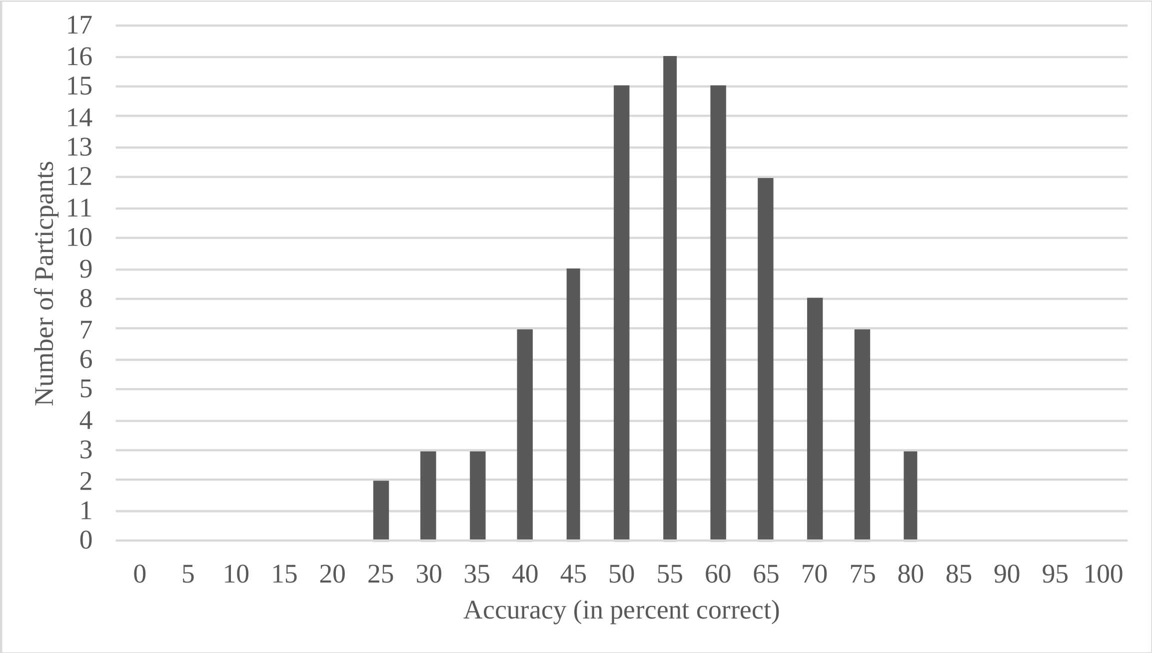


Figure 14. Distribution of participant accuracy for Evaluation 2.

4.2 Display Design and Domain Mapping

Analysis found significant main effects of display design as well as domain mapping on mean listener evaluation scores. Display conditions designed by consulting with two sonification experts with backgrounds in musical composition resulted in higher mean evaluation scores compared to the conditions that arbitrarily mapped acoustic parameters to the domain variables. When we consider the mean scores of expert display design (M = 10.61) and the arbitrarily mapped design (M = 9.42) we get a mean score difference of 1.19 between conditions. Participants that received displays with acoustic parameters mapped to variables in the weather domain resulted in higher mean evaluation scores (M = 10.42) than participants in groups that mapped acoustic parameters to variables in the health domain (M = 9.61). This results in a mean score difference of 0.81 between the two conditions. This may not seem practically significant when considering the impact of a single point or less on an evaluation score, however these 0.81 and 1.19 point differences are the equivalent to 4.1% and 6% differences in accuracy on trend identification tasks. If we consider the confidence intervals around the mean scores for both domain and display conditions the greatest possible difference in scores could be 1.95 and 2.33 points respectively. This would be the equivalent of up to 9.8% and 11.7% differences in accuracy on trend identification tasks.

The mean Evaluation 1 score is significantly higher for the expert designed weather display group compared to the arbitrarily mapped health display group. As discussed in Section 2.1.3.1 the acoustic parameters used in this study were originally selected because they were the best fit for concepts in the weather domain. For this reason, it was expected that participants in the expert designed weather display group would have the best evaluation scores, and in contrast participants in the arbitrarily mapped health display group would have the lowest performance scores. This is congruent with Hypothesis 2a and Hypothesis 2b, however there was not a statistically significant interaction between the display design and domain mapping conditions so Hypothesis 2c is rejected. The difference between mean scores for the expert designed weather display and the arbitrarily mapped health display groups is no longer significant in Evaluation 2. This suggests that the differences between display design and domain mapping conditions are mitigated by practice.

4.3 Practice with the Sonification

There was a significant main effect of practice when comparing mean scores between the first and second Evaluations, which supports Hypothesis 3. Before practice, mean evaluation score was 8.9 (SD = 2.65) and after around 45 minutes of practice for each participant there was a mean evaluation score of 11.13 (SD = 2.52). This is a mean difference of 2.23 points, or an 11.15% increase in mean listener accuracy. This is congruent with past findings of Walker and Nees (2005; see also Smith & Walker, 2005) that showed practice with feedback results in better listener performance on a point estimation task. The results of this dissertation extend those findings to show that practice with feedback also improves listener performance on trend identification tasks. Both point estimation and trend identification fall into the category of analytical listening tasks when using sonifications.

4.4 Motivation Check

As shown in Figure 12, mean scores on the Intrinsic Motivation Inventory (IMI) changed somewhat over time, but stayed relatively high overall. There was a statistically significant difference in mean IMI scores across all three time-points. This indicates that mean IMI score went up between the first and second motivation checks, which shows that participants were motivated during the practice phase of the study. But then mean IMI score decreases when comparing practice phase motivation to motivation for the second evaluation. While this may indicate that participants were a little less motivated at the end of the final evaluation, possibly due to fatigue, the mean IMI score at the third motivation check was still higher than the mean IMI score at the first motivation check. As a way to monitor overall participant motivation during the difficult task of learning and utilizing a complex sonification, it seems that participants kept their motivation throughout the study.

4.5 Models for Predicting Listener Performance

Two multiple linear regression models were created to predict user performance on the trend identification tasks. The first model was created to predict participant performance on the first evaluation. This initial evaluation tests listeners ability to utilize the sonifications without having any time to practice with the display. The first model accounts for 12% of the variance in Evaluation 1 scores and is based on participants' Musical Sophistication Index (MSI) score, Intrinsic Motivation Index (IMI) score during the first evaluation, and listening discrimination task score. Both MSI and IMI scores are significant predictors in this model. This is only a low-to-moderate predictor of overall success on the trend identification task, but it indicates that knowledge of a listener's musical sophistication and their level of motivation on the task allows us some ability to predict how well the listener will be able to perform in a scenario where they have not had time to practice with the sonification.

The second model was designed to predict listener performance on the second evaluation, which occurs after participants are given time to practice. This model was based on MSI score, motivation (IMI) score during the practice phase, motivation (IMI) score during the second evaluation task, and listener discrimination task score. This model accounts for 24% of the variance in listener scores in Evaluation 2. Musical Sophistication Index scores and listening discrimination task scores are both significant predictors in this model. These results demonstrate that measuring a listener's MSI score and their listening discrimination ability, can provide a moderate amount of prediction on how well the listener will perform, after having time to practice, on a trend identification task with five variables presented simultaneously in a sonification. The model for predicting variance in Evaluation 2 scores using listening discrimination scores and MSI scores supports Hypothesis 4.

4.5.1 Listener Discrimination Task Scores

Scores on the discrimination task are a logical predictor of final performance because the task was designed to test listeners' ability to detect changes between variables and streams that use many of the same acoustic parameters as the trend identification sonifications used in the evaluation and practice tasks. Participants with higher scores on the listening discrimination task are expected to be more sensitive to changes in acoustic stimuli, even as some of these stimuli are presented simultaneously. This was not a significant predictor of Evaluation 1 scores, but was a significant predictor for how well listeners will perform on Evaluation 2. It is possible that greater sensitivity to changes in the acoustic parameters aids listeners during the practice phase, perhaps because they are better suited to listen for target parameters after receiving feedback.

Table 11. Percentage Correct Summary for Listening Discrimination Task

	Acoustic Properties of Stimuli	Changing Parameter	% Correct
1A	Freq (400hz)	None	79
1B	Freq (500hz)	Frequency	97
2A	Freq (500hz); Tremolo (4.45hz)	None	81
2B	Freq (500hz); Tremolo (6hz)	Tremolo	98
3A	Freq (500hz); Tremolo (4.45hz); Pink Noise	None	78
3B	Freq (400hz); Tremolo (4.45hz); Pink Noise	Frequency	100
4A	Freq (500hz); Tremolo (6hz); Pink Noise	None	81
4B	Freq (500hz); Tremolo (4.45hz); Pink Noise	Tremolo speed	97
5A	Freq (500hz); Tremolo (4.45hz); Pink Noise (w/ high-pass filter)	None	81
5B	Freq (500hz); Tremolo (4.45hz); Pink Noise	High-pass filter on pink noise	94
6A	Freq (500hz); Tremolo (4.45hz); Pink Noise	None	76
6B	Freq (500hz); Tremolo (4.45hz); Pink Noise (w/ less intensity)	Decrease in pink noise intensity	87
7A	Freq (500hz); Tremolo (4.45hz); Pink Noise (w/ high-pass filter)	None	82
7B	Freq (500hz); Tremolo (6hz); Pink Noise (w/ high-pass filter)	Tremolo speed	99
8A	Freq (500hz); Tremolo (6hz); Pink Noise (w/ high-pass filter)	None	76
8B	Freq (500hz) - less intensity; Tremolo (6hz); Pink Noise (w/ high-pass filter)	Decrease in intensity of 500hz wave	60
9A	Freq (500hz); Tremolo (6hz); Pink Noise; Freq (400hz) w/ low intensity	None	76
9B	Freq (500hz); Tremolo (6hz); Pink Noise; Freq (400hz) w/ higher intensity	Intensity change on 400hz wave	39
10A	Freq (500hz); Tremolo (4.45hz); Pink Noise; Freq (400hz)	None	79
10B	Freq (500hz); Tremolo (6hz); Pink Noise; Freq (400hz)	Tremolo speed	42
11A	Freq (500hz); Tremolo (6hz); Pink Noise; Freq (400hz)	None	88
11B	Freq (500hz) - less intense; Tremolo (6hz); Pink Noise; Freq (400hz)	Lower intensity on 500hz wave	47
12A	Freq (500hz); Tremolo (6hz); Pink Noise - less intensity; Freq (400hz)	None	91
12B	Freq (500hz); Tremolo (6hz); Pink Noise; Freq (400hz)	Increased intensity of pink noise	19
13A	Freq (500hz); Tremolo (6hz); Pink Noise - less intensity; Freq (400hz)	None	87
13B	Freq (500hz); Tremolo (6hz); Pink Noise - less intensity & high-pass filter; Freq (400hz)	High-pass filter on pink noise	42

A breakdown of the stimuli for the Listening Discrimination Task is included in Table 11. A distribution of participants' scores for these stimuli groups is depicted in Figure 15. There is a trend in correct participant responses for stimuli 1 through 7 which shows slightly better performance for correctly identifying when stimuli were different (M = 96) as opposed to identifying when they were the same (M = 79.7). This trend reverses for stimuli 8 through 13, where we see performance for correctly identifying when the stimuli were different (M = 41.5) dropping below correct identification of when the stimuli stayed the same (M = 82.8).

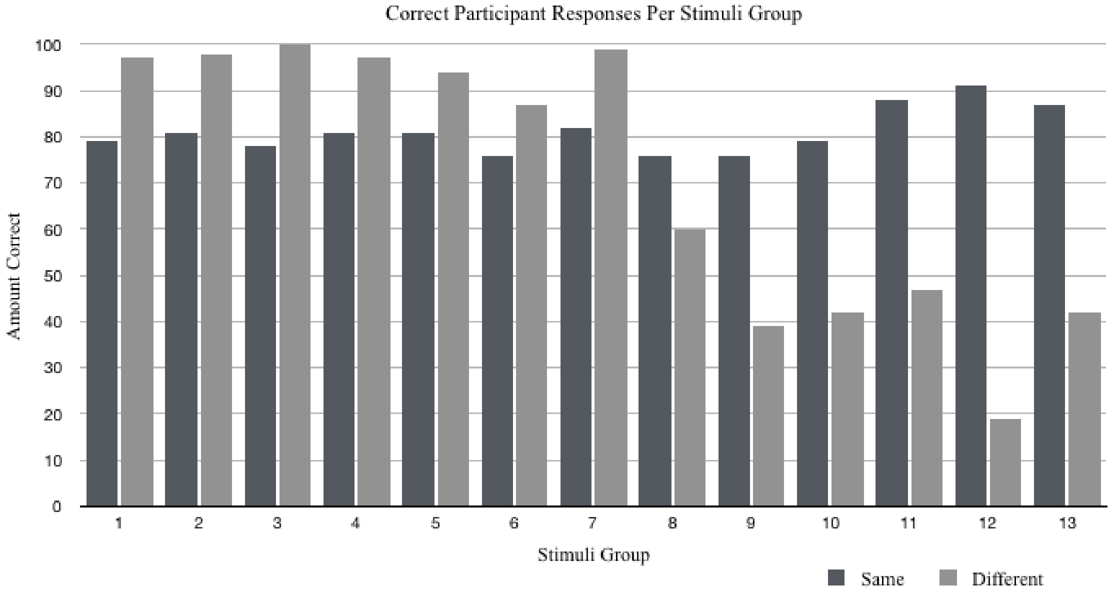


Figure 15. Distribution of participant scores for Listening Discrimination Task stimuli.

The stimuli for this task range from 1 to 5 acoustic parameters, with the intended difficulty for the stimuli becoming greater as the number of acoustic parameters increases. This is because more parameters present increases the likelihood of multiple auditory streams. Participants are then listening to determine of one of the parameters along one of the auditory streams changes.

Stimuli trials were randomized and participants received two versions of each stimuli that ether included the change (Track A followed by Track B) or as a control no change (Track A followed by Track A). Of the stimuli that participants struggled the most with, 8, 9, and 11 included a change to the intensity of the 500hz tone while the other acoustic parameters remained constant. Stimuli 10 included a change from 4.5hz to 6hz in tremolo. Stimuli 13 added a high-pass filter to the pink noise parameter, this same parameter manipulation occurs in stimuli 5, but participants had much less difficulty because fewer acoustic parameters were presented. Participants struggled the most with correctly identifying a change in stimuli 12, which was a change in intensity of the pink noise. A similar manipulation was used in stimuli 6 but with fewer parameters present participants did not have as much difficulty.

The stimuli in this task were designed based on similar acoustic parameters used in the study sonifications. A difference between the two tasks is that changes in parameters for the Listening Discrimination Task occur between two tracks that are presented to the listener in each trial. When one of the parameters changed or differed between tracks, the change was static. For instance, a change in tremolo for stimuli 10 was presented at 4.5hz in track A and then at 6hz in track B, but it was not a sweeping or dynamic change like it in the sonifications. This is because the Listening Discrimination Task was designed as a check on listener’s ability to detect a small increment between two parameter settings as a predictor of overall performance when using the sonifications later in the study.

4.5.2 Musical Sophistication Index Scores

As discussed in Section 1.5, there is some discrepancy about the correlation between past musical experience and listeners' performance on sonification tasks. Possibly due to the self-selecting tendency of people with musical aptitude to become musicians (Watson & Kidd, 1994), or because there is too much variety in how musical experience is measured across studies (Walker & Nees, 2011). This dissertation opted to measure musical experience using the Goldsmiths Musical Sophistication Index (MSI), because it includes multiple aspects of musical sophistication in addition to the usual questions about amount of past musical training and experience. This index has been validated as a sensitive metric for both musicians and non-musicians (Müllensiefen et al., 2014a; 2014b).

Participants' scores on the General Sophistication MSI were a significant predictor variable for listener performance on both the pre-test and post-test evaluations, which provides additional evidence for the effect of individual differences related to musical experience on listener performance in sonification based tasks. This is perhaps a more robust finding compared to many of the past empirical studies discussed in Section 1.5, because this dissertation implemented a pre-validated metric of musical sophistication that provides a measure across all participants in the study, rather than using a musical experience rating that is only applicable to participants who play musical instruments, such as "how many years have you played?" This finding does not completely address the correlation of past musical experience to sonification performance, but it provides additional insight on the various individual differences that effect listeners' ability to utilize complex sonifications.

CHAPTER 5. CONCLUSIONS

This dissertation investigated the necessity for expert display design and the effect of domain mapping within a sonification used for identifying trends for multiple variables simultaneously. Results of this experiment showed that expert display design and acoustic parameters specifically selected for the intended display domain lead to greater listener accuracy when users did not have a chance to practice with the auditory display before using it. However, when users are given time to practice with the display, they can learn to perform equally well with an arbitrarily mapped sonification using acoustic parameters chosen for a different context. This finding highlights the overall importance of practice when listeners are expected to utilize a multi-stream sonification for analytical listening tasks like trend identification.

In specific scenarios where a display designer wants users to utilize a multi-variable sonification without much practice, it is important that consideration is given to which acoustic parameters are used, and to how these parameters are mapped to specific domain variables within an intended context. When these sonifications are intended for a more generalized use case, such as software used in a classroom to present a variety of data domains to students with visual impairments, the display designer should focus on acoustic parameters that are optimized across auditory stream formations. Users in this type of scenario can then practice with the generalized sonification program to improve trend identification accuracy.

One limitation of the experiment in this dissertation was the amount of extended practice that listeners received. A follow-up to this study could investigate the effect of practice over additional sessions, as well as over an extended period of time. It is possible that listener accuracy was only beginning to improve and could continue to increase over subsequent practice sessions. It is also possible that even more variables could be utilized by listeners once they are able to reliably master the current five-variable sonifications.

An additional contribution of this dissertation are the models for predicting users' accuracy for complex trend identification tasks both before and after practice. The Musical Sophistication Index was a significant predictor for both pre-test and post-test performance and may be a reliable measure of individual listener differences in other auditory display tasks as well. Future sonification research should continue to employ the Goldsmiths Musical Sophistication Index as a pre-validated and standardized metric across studies as a way to better comprehend the differences between listeners. This may provide additional guidance on how to optimize display design for intended user populations.

This dissertation has shown that we are able to measure listeners' musical sophistication, motivation, and ability to discriminate between acoustic parameters, and then use those scores to predict performance on an analytical listening task. These metrics take less than 15 minutes to measure but can provide valuable insight on how long it may take some listeners to learn how to utilize complex sonifications. As we continue to design sonifications and other auditory displays for scenarios like accessibility in classroom environments, this insight is valuable in determining how to best spend time and resources for training both students and teachers.

APPENDIX A. LISTENING DISCRIMINATION TASK STIMULI

1A	Freq (400hz)
1B	Freq (500hz)
2A	Freq (500hz); Tremolo (4.45hz)
2B	Freq (500hz); Tremolo (6hz)
3A	Freq (500hz); Tremolo (4.45hz); Pink Noise
3B	Freq (400hz); Tremolo (4.45hz); Pink Noise
4A	Freq (500hz); Tremolo (6hz); Pink Noise
4B	Freq (500hz); Tremolo (4.45hz); Pink Noise
5A	Freq (500hz); Tremolo (4.45hz); Pink Noise (w/ highpass filter)
5B	Freq (500hz); Tremolo (4.45hz); Pink Noise
6A	Freq (500hz); Tremolo (4.45hz); Pink Noise
6B	Freq (500hz); Tremolo (4.45hz); Pink Noise (w/ less intensity)
7A	Freq (500hz); Tremolo (4.45hz); Pink Noise (w/ highpass filter)
7B	Freq (500hz); Tremolo (6hz); Pink Noise (w/ highpass filter)
8A	Freq (500hz); Tremolo (6hz); Pink Noise (w/ highpass filter)
8B	Freq (500hz) - less intensity; Tremolo (6hz); Pink Noise (w/ highpass filter)
9A	Freq (500hz); Tremolo (6hz); Pink Noise; Freq (400hz) w/ low intensity
9B	Freq (500hz); Tremolo (6hz); Pink Noise; Freq (400hz) w/ higher intensity
10A	Freq (500hz); Tremolo (4.45hz); Pink Noise; Freq (400hz)
10B	Freq (500hz); Tremolo (6hz); Pink Noise; Freq (400hz)
11A	Freq (500hz); Tremolo (6hz); Pink Noise; Freq (400hz)
11B	Freq (500hz) - less intense; Tremolo (6hz); Pink Noise; Freq (400hz)
12A	Freq (500hz); Tremolo (6hz); Pink Noise - less intensity; Freq (400hz)
12B	Freq (500hz); Tremolo (6hz); Pink Noise; Freq (400hz)
13A	Freq (500hz); Tremolo (6hz); Pink Noise - less intensity; Freq (400hz)
13B	Freq (500hz); Tremolo (6hz); Pink Noise - less intensity & highpass filter; Freq (400hz)

APPENDIX B. GOLDSMITHS MUSICAL SOPHISTICATION INDEX - GENERAL FACTORS

Musical Sophistication Assessment

Please respond to the following items by circling a number between 1 and 7 on the scale provided *or* fill in the blank when one is provided.

Completely Disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat Agree	Agree	Completely Agree
1	2	3	4	5	6	7

1.) I spend a lot of my free time doing music-related activities.
1 2 3 4 5 6 7

2.) I enjoy writing about music, for example on blogs and forums.
1 2 3 4 5 6 7

3.) If somebody starts singing a song I don't know, I can usually join in.
1 2 3 4 5 6 7

4.) I can sing or play music from memory.
1 2 3 4 5 6 7

5.) I am able to hit the right notes when I sing along with a recording.
1 2 3 4 5 6 7

6.) I can compare and discuss differences between two performances or versions of the same piece of music.
1 2 3 4 5 6 7

7.) I have never been complimented for my talents as a musical performer.
1 2 3 4 5 6 7

8.) I often read or search the internet for things related to music.
1 2 3 4 5 6 7

9.) I am not able to sing in harmony when somebody is singing a familiar tune.
1 2 3 4 5 6 7

10.) I am able to identify what is special about a given musical piece.
1 2 3 4 5 6 7

11.) When I sing, I have no idea whether I'm in tune or not.
1 2 3 4 5 6 7

12.) Music is kind of an addiction for me - I couldn't live without it.
1 2 3 4 5 6 7

13.) I don’t like singing in public because I’m afraid that I would sing wrong notes.
1 2 3 4 5 6 7

14.) I would not consider myself a musician.
1 2 3 4 5 6 7

15.) After hearing a new song two or three times, I can usually sing it by myself.
1 2 3 4 5 6 7

16.) I engaged in regular, daily practice of a musical instrument (including voice) for _____ years.

17.) At the peak of my interest, I practiced _____ hours per day on my primary instrument.

18.) I can play _____ musical instruments.

APPENDIX C. INTRINSIC MOTIVATION INVENTORY (IMI)

1	2	3	4	5	6	7
(not at all true)			(somewhat true)			(very true)

---

1. While I was working on the task I was thinking about how much I enjoyed it.

1 2 3 4 5 6 7

2. I did not feel at all nervous about doing the task.

1 2 3 4 5 6 7

3. I put a lot of effort into this.

1 2 3 4 5 6 7

4. I think I am pretty good at this task.

1 2 3 4 5 6 7

5. I found the task very interesting.

1 2 3 4 5 6 7

6. I felt tense while doing the task.

1 2 3 4 5 6 7

7. I think I did pretty well at this activity, compared to other people.

1 2 3 4 5 6 7

8. Doing the task was fun.

1 2 3 4 5 6 7

9. I felt relaxed while doing the task.

1 2 3 4 5 6 7

10. I enjoyed doing the task very much.

1 2 3 4 5 6 7

11. I did not try very hard to do well at this activity.

1 2 3 4 5 6 7

12. I am satisfied with my performance at this task.

1 2 3 4 5 6 7

13. I was anxious while doing the task.

1 2 3 4 5 6 7

14. I thought the task was very boring.

1 2 3 4 5 6 7

15. I tried very hard on this activity.

1 2 3 4 5 6 7

16. I felt pretty skilled at this task.

1 2 3 4 5 6 7

17. I thought the task was very interesting.

1 2 3 4 5 6 7

18. I felt pressure while doing the task.

1 2 3 4 5 6 7

19. It was important to me to do well at this task.

1 2 3 4 5 6 7

20. I would describe the task as very enjoyable.

1 2 3 4 5 6 7

21. I did not put much energy into this.

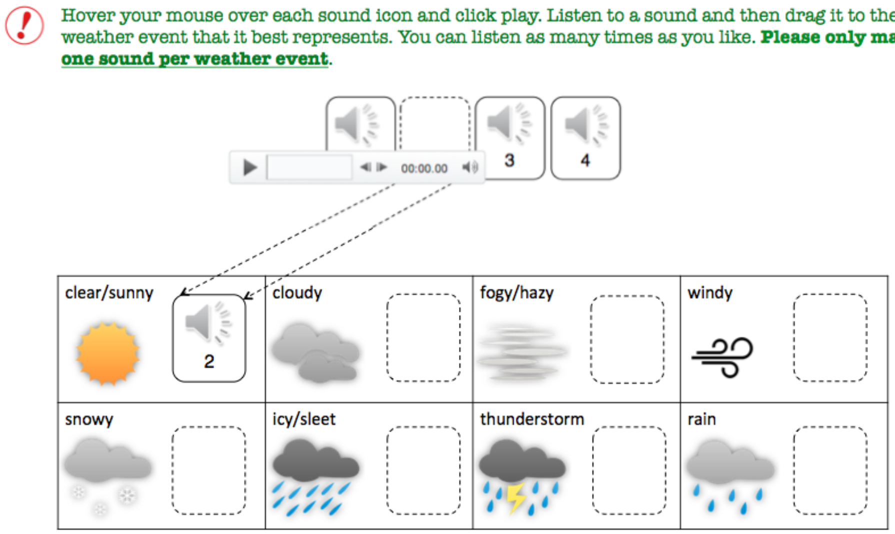
1 2 3 4 5 6 7

22. After working at this task for a while, I felt pretty competent.

1 2 3 4 5 6 7

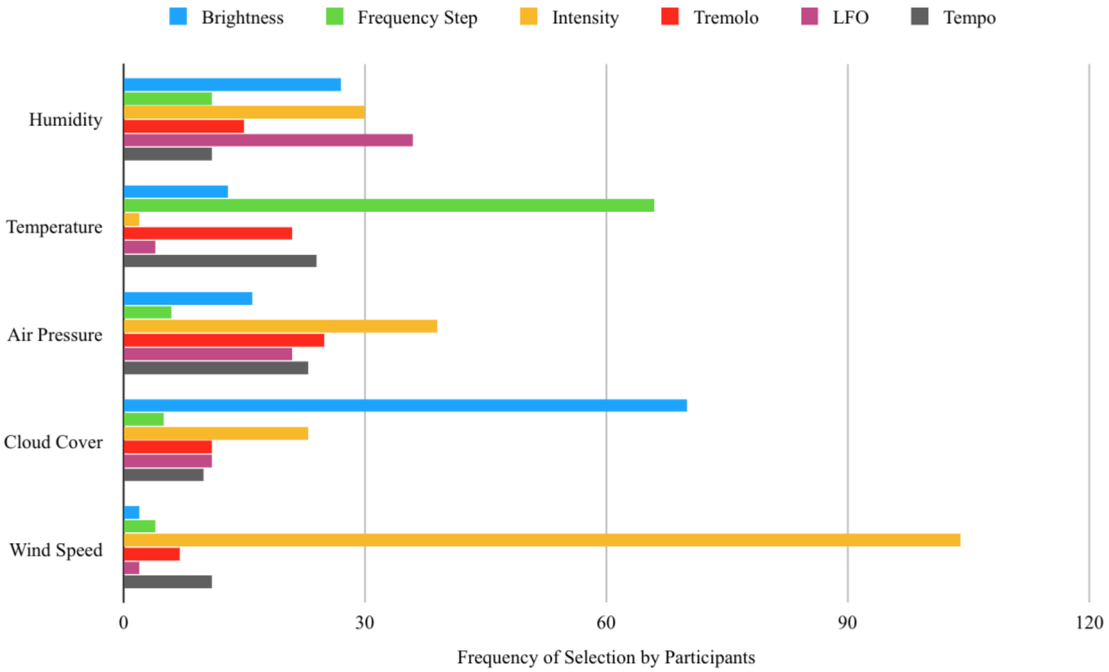
APPENDIX D. WEATHER MAPPING LISTENER RESULTS

A.1 Sound Mapping Task



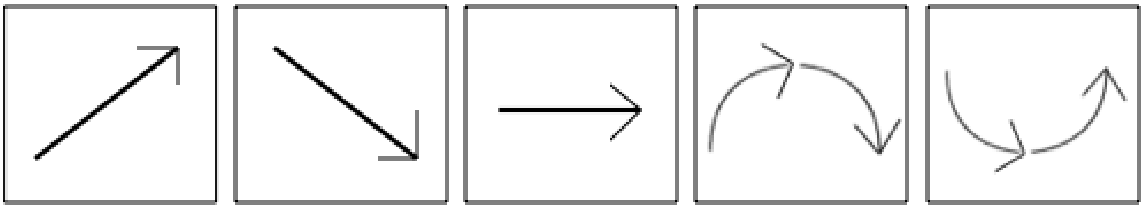
Hover your mouse over each sound icon and click play. Listen to a sound and then drag it to the weather event that it best represents. You can listen as many times as you like. **Please only match one sound per weather event**.

A.2 Results of Listener Mapping

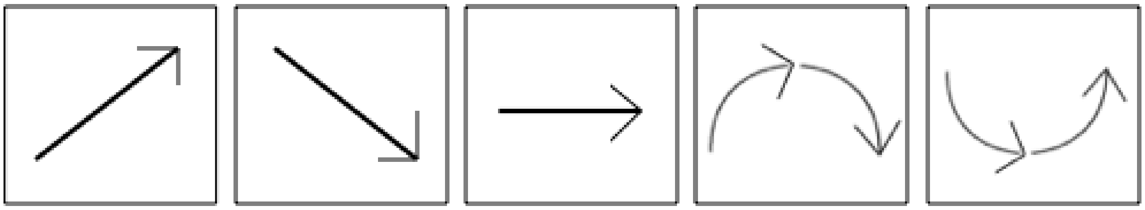


APPENDIX E. EXAMPLE EVALUATION ANSWER SHEET

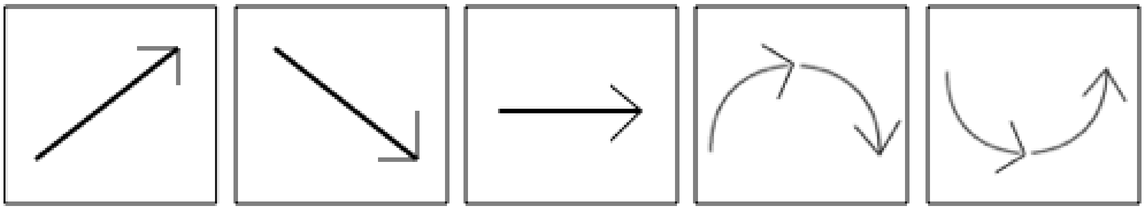
1.



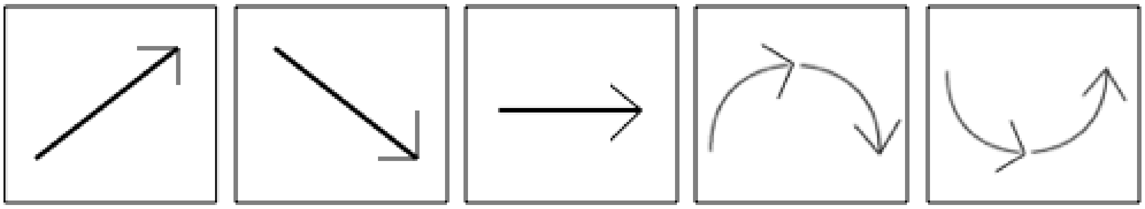
2.



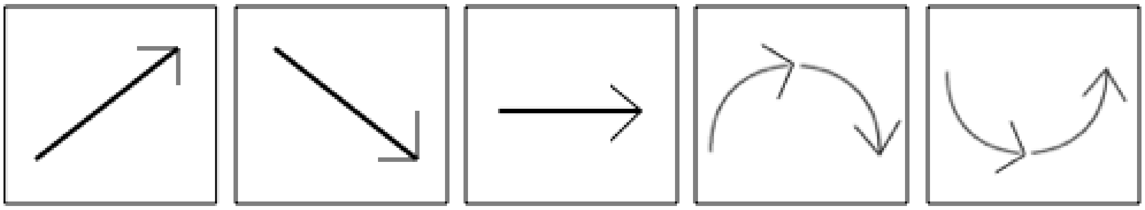
3.



4.



5.



REFERENCES

Anderson, J., & Sanderson, P. (2004). Designing sonification for effective attentional control in complex work domains. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48*(16), 1818-1822.

Anderson, J. E., & Sanderson, P. (2009). Sonification design for complex work domains: dimensions and distractors. *Journal of Experimental Psychology: Applied*, *15*(3), 183.

Anstis, S. M., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance, 11*(3), 257.

Barrass, S. (1997). *Auditory information design*. Australian National University.

Barrass, S., & Kramer, G. (1999). Using sonification. *Multimedia Systems*, *7*(1), 23-31.

Barrass, S. (2005). A perceptual framework for the auditory display of scientific data. *ACM Transactions on Applied Perception, 2*(4), 389-402.

Beauvois, M. W., & Meddis, R. (1996). Computer simulation of auditory stream segregation in alternating-tone sequences. *The Journal of the Acoustical Society of America, 99*(4), 2270-2280.

Bly, S. (1994). Multivariate data mappings. In G. Kramer (Ed.), *Auditory Display: Sonification, audification, and auditory interfaces.* (405-416). Reading, MA: Addison-Wesley Publishing Company.

Bonebright, T. L, Nees, M. A., Connerley, T. T., & McCain, G. R. (2001). Testing the effectiveness of sonified graphs for education: A programmatic research project. *Proceedings of International Conference on Auditory Display*. Espoo, Finland.

Bregman, A. S. (1990). Auditory scene analysis: The perceptual organization of sound. MIT press.

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89*(2), 244.

Bregman, A. S., & Liao, C. (1984). Stream segregation based on fundamental frequency and spectral peak II: Effects of fixed spectral window and local peak shaping through additive synthesis. Unpublished manuscript, Psychology Department, McGill University.

Bregman, A. S., Liao, C., & Levitan, R. (1990). Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology/Revue canadienne de psychologie, 44*(3), 400.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology/Revue canadienne de psychologie, 32*(1), 19.

Brewster, S. A. (1994). Providing a structured method for integrating non-speech audio into human-computer interfaces. *Doctoral dissertation*. University of York.

Brewster, S. A., Wright, P. C., & Edwards, A. D. (1994). A detailed investigation into the effectiveness of earcons. In G. Kramer (Ed.), *Auditory Display: Sonification, audification, and auditory interfaces.* (471-498). Reading, MA: Addison-Wesley Publishing Company

Brochard, R., Drake, C., Botte, M. C., & McAdams, S. (1999). Perceptual organization of complex auditory sequences: effect of number of simultaneous subsequences and frequency separation. Journal of Experimental Psychology: Human Perception and Performance, 25(6), 1742.

Brown, L. M., & Brewster, S. A. (2003). Drawing by ear: Interpreting sonified line graphs. *Proceedings of International Conference on Auditory Display*. Boston, MA.

Brown, L. Brewster, S., Ramloll, R., Yu, W., & Riedel, B. (2002). Browsing modes for exploring sonified line graphs. *Presented at 16th British HCI Conference*. London, UK.

Chapman, D. W. (1932). Relative Effects of Determinate and Indeterminate" Aufgaben". *The American Journal of Psychology*, 163-174.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975-979.

Cusack, R., Decks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance, 30*(4), 643.

Cusack, R., & Roberts, B. (2000). Effects of differences in timbre on sequential grouping. *Perception & Psychophysics*, *62*(5), 1112-1120.

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. (1994). Facilitating internalization: The self- determination theory perspective. *Journal of Personality, 62*, 119-142.

Deutsch, D. (1974). An auditory illusion. *The Journal of the Acoustical Society of America, 55*(S1), S18-S19.

Deutsch, D. (1975). Two-channel listening to musical scales. *The Journal of the Acoustical Society of America, 57*(5), 1156-1160.

Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive Psychology*, *5*(3), 322-337.

Drake, C., & Botte, M. C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics, 54,* 277-286.

Fitch, W. T., & Kramer, G. (1994). Sonifying the body electric: Superiority of an auditory over visual display in a complex, multivariate system. In G. Kramer (Ed.), *Auditory Display: Sonification, audification, and auditory interfaces.* (307-325). Reading, MA: Addison-Wesley Publishing Company.

Flowers, J. H. (2005). Thirteen years of reflection on auditory graphing: Promises, pitfalls, and potential new directions. *Faculty Publications*, Department of Psychology, 430.

Flowers, J. H., Buhman, D. C., & Turnage, K. D. (1997). Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 39*(3), 341-351.

Flowers, J. H., Whitwer, L. E., Grafel, D. C., & Kotan, C. A. (2001). Sonification of daily weather records: Issues of perception, attention and memory in design choices. Faculty Publications, Department of Psychology. Paper 432. http://digitalcommons.unl.edu/psychfacpub/432.

Frysinger, S. P. (1990, August). Applied research in auditory data representation. In Electronic Imaging'90, Santa Clara, 11-16 Feb'102 (pp. 130-139). International Society for Optics and Photonics.

Gaver, W. W., Smith, R. B., & O'Shea, T. (1991, April). Effective sounds in complex systems: The ARKola simulation. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems* (85-90). ACM.

Gordon, J. W., & Grey, J. M. (1978). Perception of spectral modifications on orchestral instrument tones. *Computer Music Journal*, 24-31.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America, 61*(5), 1270-1277.

Grey, J. M., & Moorer, J. A. (1977). Perceptual evaluations of synthesized musical instrument tones. *The Journal of the Acoustical Society of America, 62*(2), 454-462.

Hartmann, W. M., & Johnson, D. (1991). Stream segregation and peripheral channeling. *Music Perception*, 155-183.

Hayward, C. (1994). Listening to the earth sing. In G. Kramer (Ed.), *Auditory Display* (369-404). Reading, MA: Addison-Wesley Publishing Company.

Heise, G. A., & Miller, G. A. (1951). An experimental study of auditory patterns. *The American Journal of Psychology, 64*(1), 68-77.

Herman, T. Drees, J. M., & Ritter, H. (2003). Broadcasting auditory weather reports – A pilot project. *Proceedings of the 2003 International Conference on Auditory Display (ICAD),* 6-9 July 2003, Boston MA.

Hirsh, I. J. (1959). Auditory perception of temporal order. *The Journal of the Acoustical Society of America, 31*(6), 759-767.

Houtsma, A. J. (1997). Pitch and timbre: Definition, meaning and use. *Journal of New Music Research, 26*(2), 104-115.

Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception*, 361-382.

Idson, W. L., & Massaro, D. W. (1976). Cross-octave masking of single tones and musical sequences: The effects of structure on auditory recognition. *Perception & Psychophysics*, *19*(2), 155-175.

Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., & Klapp, S. T. (1995). Test of attentional flexibility in polyrhythmic patterns. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 293-307.

Kidd, G. R., & Watson, C. S. (1987). The perception of complex multidimensional sounds. *Journal of Acoustic Society of America, 81*(1).

Kramer, G. (1994a). Auditory display: Sonification, audification, and auditory interfaces. Perseus Publishing.

Kramer, G. (1994b). An introduction to auditory display. In G. Kramer (Ed.), *Auditory Display: Sonification, audification, and auditory interfaces.* (1-77). Reading, MA: Addison-Wesley Publishing Company.

Kramer, G. (1994c). Some organizing principles for representing data with sound. In G. Kramer (Ed.), *Auditory Display: Sonification, audification, and auditory interfaces.* (185-221). Reading, MA: Addison-Wesley Publishing Company.

Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J., Miner, N., & Neuhoff, J. (2010). Sonification report: Status of the field and research agenda. *Faculty Publication*, Department of Psychology. Paper 444.

Lacherez, P., Seah, E. L., & Sanderson, P. (2007). Overlapping melodic alarms are almost indiscriminable. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(4), 637-645.

Loeb, R. G., & Fitch, W. T. (2002). A laboratory evaluation of an auditory display designed to enhance intraoperative monitoring. *Anesthesia & Analgesia*, 94(2), 362-368.

Massaro, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, *79*(2), 124.

McAdams, S., Beauchamp, J. W., & Meneguzzi, S. (1999). Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters. *The Journal of the Acoustical Society of America, 105*(2), 882-897.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research, 58*(3), 177-192.

McCabe, S. L., & Denham, M. J. (1997). A model of auditory streaming. *The Journal of the Acoustical Society of America, 101*(3), 1611-1621.

McGookin, D. K., & Brewster, S. A. (2004). Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition. *ACM Transactions on Applied Perception, 1,* 130–155.

Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, *48*(2), 169-178.

Mezrich, J. J., Frysinger, S., & Slivjanovski, R. (1984). Dynamic representation of multivariate time series data. *Journal of the American Statistical Association, 79*(385), 34-40.

Miller, G. A., & Heise, G. A. (1950). The trill threshold. *The Journal of the Acoustical Society of America*. *22*(5), 637-638.

Moore, B. C., & Gockel, H. (2002). Factors influencing sequential stream segregation. *Acta Acustica United with Acustica, 88*(3), 320-333.

Müllensiefen, D., Gingras, B., Stweard, L., & Musil, J. (2014a) The musicality of non-musicians: An index for measuring musical sophistication in the general population. PLoS ONE 9(2): e89642

Müllensiefen, D., Gingras, B., Stweard, L., & Musil, J. (2014b) The goldsmiths musical sophistication index (Gold-MSI): *Technical Report and Documentation v.10*. London: Goldsmiths, University of London.

Nees, M. A., & Walker, B. N. (2007). Listener, task, and auditory graph: Toward a conceptual model of auditory graph comprehension. *Proceedings of the International Conference on Auditory Display (ICAD2007),* Montreal, Canada. 266-273.

Nees, M. A., & Walker, B. N. (2008). Data Density and Trend Reversals in Auditory Graphs: Effects on Point Estimation and Trend Identification Tasks. *ACM Transactions on Applied Perception, 5*(3) 1-24.

Neuhoff, J. G., Wayand, J., & Kramer, G. (2002). Pitch and loudness interact in auditory displays: Can the data get lost in the map? *Journal of Experimental Psychology: Applied, 8*(1), 17-25.

Neuhoff, J. G., & McBeath, M. K. (1996). The Doppler illusion: the influence of dynamic intensity change on perceived pitch. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 970-985.

Neuhoff, J. G., McBeath, M. K., & Wanzie, W. C. (1999). Dynamic frequency change influences loudness perception: a central, analytic process. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 1050-1059.

Norman, D. A. (1967). Temporal confusions and limited capacity processors. *Acta Psychologica, 27,* 293-297.

Plant, R. W., & Ryan, R. M. (1985). Intrinsic motivation and the effects of self-consciousness, self- awareness, and ego-involvement: An investigation of internally-controlling styles. Journal of Personality, 53, 435-449.

Pollack, I. (1952). The information of elementary auditory displays. *The Journal of the Acoustical Society of America, 24*(6), 745-749.

Pollack, I. (1953a). Assimilation of sequentially encoded information. *The American Journal of Psychology*, 421-435.

Pollack, I. (1953b). The information of elementary auditory displays. II. *The Journal of the Acoustical Society of America*, *25*(4), 765-769.

Polli, A. (2004). Atmospherics/weather works: A multi- channel storm sonification project. Proceedings of the 2004 International Conference on Auditory Display (ICAD), 6-9 July 2004, Sydney Australia.

Rose, M. M., & Moore, B. C. (2000). Effects of frequency and level on auditory stream segregation. *The Journal of the Acoustical Society of America, 108*(3), 1209-1214.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. Journal of Personality and Social Psychology, 43, 450-461.

Ryan, R. M., Mims, V., & Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. Journal of Personality and Social Psychology, 45, 736-750.

Sanderson, P. M., Watson, M. O., & Russell, W. J. (2005). Advanced patient monitoring displays: tools for continuous informing. *Anesthesia & Analgesia, 101*(1), 161-168.

Sanderson, P. M., Wee, A., & Lacherez, P. (2006). Learnability and discriminability of melodic medical equipment alarms. *Anaesthesia*, *61*(2), 142-147.

Scaletti, C. (1994). Sound Synthesis Algorithms for Auditory Data Representations. In G. Kramer (Ed.), *Auditory Display* (223-251). Reading, MA: Addison-Wesley Publishing Company.

Schuett, J. H. (2010). Limits on the number of concurrent auditory streams. *Masters Thesis*. James Madison University, Harrisonburg, VA.

Schuett, J. H., & Walker, B. N. (2013) Measuring comprehension in sonification tasks that have multiple data streams. *In Proceedings of the 8th Audio Mostly Conference.* ACM.

Schuett, J. H., Winton, R. J., Batterman, J. M., & Walker, B. N. (2014). Auditory weather reports: demonstrating listener comprehension of five concurrent variables. *In Proceedings of the 9th Audio Mostly: A Conference on Interaction with Sound*. ACM. 17.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379-423, 623-656.

Singh, P. G. (1987). Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre?. *The Journal of the Acoustical Society of America, 82*(3), 886-899.

Singh, P. G., & Bregman, A. S. (1997). The influence of different timbre attributes on the perceptual segregation of complex-tone sequences. *The Journal of the Acoustical Society of America, 102*(4), 1943-1952.

Smith, D. R. & Walker, B. N. (2002). Tick-marks, axes, and labels: The effects of adding context to auditory graphs. *Proceedings of the Eighth International Conference on Auditory Display ICAD2002*, Kyoto, Japan (02-05 July) pp 362-367.

Smith, D. R., & Walker, B. N. (2005). Effects of auditory context cues and training on performance of a point estimation sonification task. *Applied Cognitive Psychology, 19*(8), 1065-1087.

Song, H. J., & Beilharz, K. (2007). Concurrent auditory stream discrimination in auditory graphing. *Journal of Computers, 3*, 79-87.

Stevens, S. S. (1957). On the psychophysical law. *Psychological review*, *64*(3), 153.

Stevens, S. S. (1975). *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects*. New York: Wiley.

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12*(4), 242-248.

Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior, 3*(6), 449-459.

Treisman, A. M. (1971). Shifting attention between the ears. *The Quarterly Journal of Experimental Psychology, 23*(2), 157-167.

van Noorden, L.P.A.S. (1975). Temporal coherence in the perception of tone sequences. *Unpublished doctoral dissertation*, Eindhoven University of Technology.

Walker, B. N. (2000). *Magnitude estimation of conceptual data dimensions for use in sonification*. Doctoral Dissertation. Rice University, Houston, TX.

Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied*, *8*(4), 211.

Walker, B. N., & Ehrenstein, A. (2000). Pitch and pitch change interact in auditory displays. *Journal of Experimental Psychology: Applied*, *6*(1), 15-30.

Walker, B. N., & Kramer, G. (1996). Mappings and metaphors in auditory displays: An experimental assessment. *Proceedings of the 3rd International Conference on Auditory Display* (ICAD96). (71-74). Palo Alto, CA.

Walker, B. N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. In J. Neuhoff (Ed.) *Ecological psychoacoustics* (pp. 150-175). New York: Academic Press.

Walker, B. N., & Kramer, G. (2005). Mappings and metaphors in auditory displays: An experimental assessment. *ACM Transactions on Applied Perception (TAP), 2*(4), 407-412.

Walker, B. N., Kramer, G., & Lane, D. M. (2000). Psychophysical scaling of sonification mappings. *Proceedings of the Sixth International Conference on Auditory Display ICAD2000*, Atlanta, GA.

Walker, B. N., & Lane, D. M. (2001). Psychophysical scaling of sonification mappings: A comparison of visually impaired and sighted listeners. *Proceedings of the*

Seventh International Conference on Auditory Display ICAD2001, Espoo, Finland (28 July-01 August) pp 90-94.

Walker, B. N., & Mauney, L. M. (2010). Universal design of auditory graphs: A comparison of sonification mappings for visually impaired and sighted listeners. *ACM Transactions on Accessible Computing (TACCESS)*, *2*(3), 12.

Walker, B. N., & Nees, M. A. (2005). Brief training for performance of a point estimation sonification task. *Proceedings of the International Conference on Auditory Display (ICAD2005)*, Limerick, Ireland.

Walker, B. N., & Nees, M. A. (2011). Theory of sonification. In Hermann, T., Hunt, A., & Neuhoff, J. G. (Eds.), *The Sonification Handbook.* (9-39). Logos Publishing House, Berlin, Germany.

Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. (1969). Auditory sequence: Confusion of patterns other than speech or music. *Science*, *164*(3879), 586-587.

Watson, C. S., & Kidd, G.R. (1994). Factors in the design of effective auditory displays. *Proceedings of the Second International Conference on Auditory Display ICAD `94*, Santa Fe Institute, New Mexico.

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 45-52.

Yost, W. A. (1997). The cocktail party problem: Forty years later. In R. Timothy (Ed.), *Binaural and Spatial Hearing in Real and Virtual Environments*. (329-347). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.