MIXED SPEECH AND NON-SPEECH AUDITORY DISPLAYS: IMPACTS OF DESIGN, LEARNING, AND INDIVIDUAL DIFFERENCES IN MUSICAL ENGAGEMENT

Grace Li

Georgia Institute of Technology,
648 Cherry St NW
Atlanta, GA 30313, USA
tli.grace@gatech.edu

Bruce N. Walker

Georgia Institute of Technology,
648 Cherry St NW
Atlanta, GA 30313
bruce.walker@psych.gatech.edu

ABSTRACT

Information presented in auditory displays is often spread across multiple streams to make it easier for listeners to distinguish between different sounds and changes in multiple cues. Due to the limited resources of the auditory sense and the fact that they are often untrained compared to the visual senses, studies have tried to determine the limit to which listeners are able to monitor different auditory streams while not compromising performance in using the displays. This study investigates the difference between non-speech auditory displays, speech auditory displays, and mixed displays; and the effects of the different display designs and individual differences on performance and learnability. Results showed that practice with feedback significantly improves performance regardless of the display design and that individual differences such as active engagement in music and motivation can predict how well a listener is able to learn to use these displays. Findings of this study contribute to understanding how musical experience can be linked to usability of auditory displays, as well as the capability of humans to learn to use their auditory senses to overcome visual workload and receive important information.

1. INTRODUCTION

People regularly use visual displays to aid in monitoring data and increasing their situation awareness. With more advanced technology and research, these displays have been beneficial in helping people gather information. However, the amount of information users are able to attend to visually remains limited, even as the demand for more information increases. Researchers have turned to auditory displays as an additional channel, and have studied the benefits of audio versus visual information on peoples' ability to comprehend and retain information presented. This study focuses directly on auditory display design and the impact of listener differences such as musical experience and motivation on usability of auditory displays.

In the example of an anesthesiologist who needs to monitor a patient's vitals during surgery, visual displays can be overwhelming. These displays may prevent anesthesiologists from visually attending to other areas of their workspace. To ease the workload in the visual field, a mix of visual and auditory displays may be used together, where the auditory display would cue the anesthesiologist to look at the information on the visual display. However, in circumstances when the anesthesiologist cannot visually attend to the visual display, auditory displays may prove to be beneficial in informing the anesthesiologist on the status of their patient. Auditory displays prevent overload of information in other daily activities most people encounter, such as listening to the news or weather report in the morning while stuck in bed, or changing music playlists while driving.

The term *sonification* describes a subtype of auditory display that typically uses non-speech audio to present information by translating relationships in data into sounds that human listeners are able to comprehend [1]. Information in auditory displays is mapped to certain sounds that help listeners understand and interpret the information. Bregman and Campbell define auditory streams as a "sequence of auditory events" that are blended together to convey a message or an idea into one single "stream" [2]. These auditory sequences can be different, yet related, in order for them to fit together and present information that makes sense to the listener. These streams of sounds can include manipulations of vario*us* acoustic properties, such as pitch and tempo. Many studies have looked at the use of multi-stream auditory displays in an anesthesiologist's workstation, specifically looking at the effects of mapping multiple pieces of information to fewer auditory streams. Fitch and Kramer mapped eight different health-related variables to two different streams and found that participants improved with practice and were able to manage all the variables [3]. They concluded that auditory systems that simultaneously convey a number of variables can be more effective than visual displays, separating variables into individual pieces of information to perceive one at a time [3]. In a similar study, Loeb and Fitch used actual anesthesiologists to see if they were able to monitor six different variables at once, and found that with little practice, the clinicians were able to identify all the variables in two different streams and simultaneously decipher and respond to critical events [4]. These multi-stream auditory display studies suggest that listeners can accurately monitor up to eight different variables combined into three separate auditory streams within a complex auditory display [3], [5]. Additionally, Schuett found that participants were able to follow about five auditory variables at a time which were blended together to form three more dominant comprehensive streams [6] [7]. Applying the information to a more practical setting, Schuett created auditory displays with three auditory streams using five acoustic parameters, each representing five different variables related to weather or health, and observed participants' ability to interpret information from the auditory display [8]. For instance, one of the health-related variables was Heart Rate, which was mapped to the tempo of one of the streams, while Respiratory Rate was mapped to the frequency of that same stream. Findings suggested that participants were able to learn to comprehend the auditory display and were able to perform better with practice.

Auditory display research has also looked into the effects of individual differences in listening abilities, familiarity, and practice on the usability of these displays. Watson and Kidd suggest that listeners' perceptual and cognitive abilities play an important role in the systems' usability, while comprehension of these displays may be a result of the listeners' musical ability [9]. They propose that there must be a subjective perceptual difference among the participants when using auditory displays. This points towards musical training and experience as differences that may impact listeners' ability to understand auditory displays. Brochard, Drake, Botte, and McAdams had participants listen to three auditory streams and signal when they found small temporal irregularities within the auditory streams [10]. They found that participants who were musically trained performed better at detecting irregularities than those who were not musically trained. However, there were no significant interactions between musical training and the other variables such as frequency grouping and target location. Lacherez, Seah, and Sanderson concluded that failure in stream segregation was the limiting factor for listeners' perception, even for those who were musically trained [11]. Schuett suggests that although the link between musical experience and stream segregation is unclear, it seems to be that some form of familiarity with the acoustic properties of the auditory display may be helpful in stream segregation [8]. Walker and Nees looked into the role of training and found that practice with feedback led to significantly lower errors in point estimation tasks using sonification more so than no practice, practice only, practice with visual prompts, and conceptual training [12]. Therefore, having knowledge of the results during training can have a positive impact on listeners' improvement and performance.

Because much of the population is not familiar with sonification, there are challenges in incorporating sonification into daily activities. People are becoming more familiar with speech-based auditory displays such as the speech commands in GPSs, Siri and Alexa, which may make speech seem to be a viable alternative to sonification. In some cases, that may be the case; in other cases, not. Nevertheless, when multiple channels of data need to be conveyed, there may well be challenges that would arise with multiple streams of speech. Ericson, Brungart, and Simpson list factors that influence comprehension of speech displays in the context of air force pilots [13]. They determined that the addition of simultaneous voices would decrease the performance of the listener, so keeping the number of speech sounds to a minimum in a display would be best. Differing *characteristics* of the voices can help segregate speech sounds in a display, where pitch/frequency, speaking rate, accents, and intonation might help listeners comprehend the different streams. Finally, *spatially* separating speech sounds in a display can also help listeners comprehend each speech sound, more so than the other techniques, which would only increase intelligibility of one or two speech sounds, at the expense of losing information from the other speech sounds [13]. In a more applied situation, Simpson, Brungart, Dallman, Joffrion, Presnar, and Gilkey tested spatial audio displays in general aviation environments with trained pilots [14]. They found that spatial audio displays effectively improve pilots' situation awareness and safety in general aviation environments when used for both navigation and altitude monitoring. Similarly, Simpson, Brungart, Gilkey, and McKinley found that pilots were very accepting of the spatial audio display and showed low annoyance levels, which suggests that spatial audio displays are important to comprehend spoken information and to prevent overload or annoyance [15].

Ericson et al. have shown that speech displays can make it difficult to monitor different speech streams because speech streams tend to mask each other [13]. Multiple speech streams can be overwhelming and prevent listeners from obtaining adequate information. Methods to keep speech streams separate and intelligible are effective, but there are still limits to how many streams can be followed. Li, Tang, Hickling, Yau, Brecknell, and Sanderson found that speech cues can lead to more accurate responses in identifying information than earcons, however that may be due to the fact that people are more familiar with speech cues than earcons [16].

Even with all the focus on the use of sonification and speech displays to convey data, there has been a gap in knowledge about the effects of speech streams interactions with the sonification on listeners' ability to perceive information. Walker and Nees mention the wealth of knowledge in sonification during concurrent visual and auditory tasks, but a lack in the degree to which non-speech audio interacts with concurrent processing of other sounds such as speech [2]. The purpose of the present study is to combine the benefits of both sonification and speech auditory displays into a mixed auditory display in order to see the effects of the interaction on listeners' comprehension and performance. The mixed displays should minimize the unfamiliarity of sonification, and introduce speech, while also ensuring that there are not too many speech streams to distract or mask the other streams.

2. STUDY OVERVIEW

The study is a continuation and adaptation of Schuett's dissertation [8]. Participants assumed the role of an anesthesiologist and detected trends in body vitals of a virtual patient using an auditory display with five variables combined into three streams. There are a total of four displays variants: one that uses Schuett's [8] "Health" non-speech display; one with all speech; and two with a mixture of speech and non-speech. The speech sounds were added into the display using techniques highlighted by Ericson et al. [13] by separating them spatially, and by frequency. Participants were randomly assigned to one of the displays and took a pretest to see their initial comprehension of the display. Then they completed a practice phase with feedback, and finally, completed a final test to see if they were able to improve their comprehension of the display. The scores were compared across all the three display conditions to see which display had the highest learnability and performance. Additionally, subjective measurements through motivation surveys and musical experiences were used to assess the impact that individual differences may also have on using the auditory displays before and after practice.

3. METHODS

3.1. Participants

Participants in this study were 97 students at a U.S. university between the ages of 17 and 29 (M = 20.0, SD = 1.80), who received extra credit in a college class. There was a total of 32 to 33 participants for each of the three between-subjects conditions tested. Participants all reported normal or corrected-to-normal vision and normal hearing.

Table 1: Display Design

Condition	Context		Basis
Non-Speech	Non-speech: All health variables		Based on the judgments of the sound designers to best-fit health concepts to the acoustic parameters outlined by Schuett (2017)
2 Speech	Speech: Heart rate, Blood pressure	Non-speech: Blood oxygen level, Respiratory rate, Body Temperature	Based on the judgement of which health concepts fit best with speech streams
3 Speech	Speech: Heart rate, Blood pressure, Body Temperature	Non-speech: Blood oxygen level, Respiratory rate	Based on the judgment of which health concepts fit best with speech streams
Speech	Speech: All health variables		Based on the judgment of what sounds the best when all five speech streams are played together.

Table 2: Location and Acoustic Mappings

	Left Ear	Centered	Right Ear
Variable Location	Respiratory Rate	Body Temperature	Blood Oxygen Level
	Heart Rate	--	Blood Pressure
Acoustic Parameters	**Left Pan**	**Centered**	**Right Pan**
	Frequency (Pitch)	Chord (Intensity changes)	Pink Noise (Intensity changes)
	Tremolo (Speed)	--	Filter (filter on pink noise)

Table 3: Acoustic and Speech Parameters

Concept Variable	Acoustic Parameter	Speech Parameter
Respiratory Rate	Frequency	Numeric respiratory rate value Uses lower pitched voice.
Heart Rate	Tremolo	Numeric heart rate value Uses a higher pitched voice.
Body Temperature	Intensity (chord)	Numeric body temperature value Uses a monotone, robotic voice.
Blood Oxygen Level	Pink Noise Intensity	Numeric blood oxygen level value Uses a higher pitched voice.
Blood Pressure	Filter (on pink noise)	Numeric blood pressure value (two numbers) Uses a lower pitched voice.

3.2. Display Design and Mapping

The methods of this experiment followed closely to Schuett's dissertation [8], but, with adjusted sound files and some minor procedural changes. The purpose of the displays is to determine which auditory display mappings (speech, non-speech, and mixed) results in highest performance, by comparing their learnability to one another. There were four display mappings. Table 1 includes all four mappings, using the same health variables. One mapping was identical to the "Health" mapping in Schuett's dissertation [8], which maps five health variables, specifically those used by anesthesiologists. Another mapping used the same health variables in Schuett's dissertation [8] but introduced speech streams based on the study of speech displays by Ericson et al. [13]. The two mixed displays had a combination of speech and non-speech streams. One had two non-speech streams and three speech streams, and the other had three non-speech streams and two speech streams.

The auditory streams were separated in stereo space by panning one into the left ear, one into the right ear, and the third centered. The centered stream was used to only represent one variable, while both the left and right represented two variables combined in one stream. The use of three streams was to segregate the five variables for listeners. Table 2 shows the mapping of the health variables to their respective ears.

Non-Speech mapping. This display is identical to Schuett's "Best-fit Display Mapping (Health)" [8]. Table 2 summarizes the acoustic mapping of the sonifications for the Non-Speech display.

The data trends represented by each of these five parameters could increase, decrease, remain constant, increase-then-decrease, or decrease-then-increase over time. The display was intended to represent informative trends that any of the health parameters could have in a given time frame. The context of the health data was chosen for this condition, which is congruent with past sonification of health related concepts such as Fitch and Kramer [3] and Anderson and Sanderson [5]. Respiratory rate and heart rate were paired together in the left ear because the two are connected conceptually; and similarly, blood oxygen level and blood pressure were also paired due to their connection to one another in the human body. Body temperature is least connected to the other four variables, so it remained in its own stream in the stereo-centered location.

Mixed Displays: 2-Speech mapping and 3-Speech mapping. These displays added speech into certain variables of the display. The mixture of the two stream types incorporated findings from Fitch and Kramer and reflected the optimal design for speech auditory displays as indicated by Ericson et al. [3], [13]. For the 2-Speech display, two of the variables were mapped using speech and three of the variables were mapped using non-speech sonification. For the 3-Speech display, three of the variables were mapped using speech sounds and the other two remained non-speech. The general layout for each display was similar to the Non-Speech display, where each variable remained in its respective ears and followed its respective acoustic parameter. Table 3 includes the acoustic and speech parameters for each variable.

Speech Display. The final display had all five variables represented by five speech sounds. The speech parameters are listed in Table 3. Following pilot testing, this display was not included in the experiment due to the difficulty participants had with it. Even when intentionally listening to the display sounds, it was difficult to concentrate and monitor a single variable, let alone five speech variables.

3.3. Materials

Throughout the duration of the study, participants wore SONY MDR-V150 Headphones, sat in front of a computer in a computer lab, and completed the study via an automated Qualtrics survey. This was a slight procedural differences from Schuett's study in which participants were run one at a time and researchers were heavily involved during each step [8].

Listening Discrimination Task. The point of the Listening Discrimination Task is to see if differences in individual

performance on the task affect performance on the use of the auditory display. Individual differences allow some participants to have a "trained ear", which allows them to be better at discerning smaller differences between acoustic stimuli.

Participants' abilities were assessed separately from the main study. The Listening Discrimination Task after Schuett [8] required participants to listen to one audio track, followed by another, and determine if the first and second track were the same or different. The first and second track were either the same, or differed by one acoustic parameter each time. The task increased in difficulty when the number of acoustic parameters in the tracks increased. When there was only one acoustic parameter, a change across that single parameter was relatively easy for the listener to discern. But when there were multiple acoustic parameters in each track, detecting the presence of a change became increasingly difficult.

For each Listening Discrimination Task trial, participants were presented Track A and then Track B, and given a choice "same" or "different" to choose from. This task consisted of 26 total trials. In half of the trials, Tracks A and B were the same, and in the other half they were different. The trial difficulty was presented in a randomized order for each participant through Qualtrics. The acoustic parameters used for each of the thirteen acoustic groupings are included in Appendix A.

Intrinsic Motivation Inventory. This study also used the same Intrinsic Motivation Inventory (IMI) scale [17], [18], as Schuett's study [8], which participants completed three times throughout the study. The scale measures subjective motivation towards a specific task during the study. The first was administered after the pretest to gauge motivation during the pretest phase. The second occurred at the end of the practice with feedback phase, and the third occurred after the posttest. The purpose of these was to determine if participants got bored or tired throughout the study and if it would have an effect on the participant responses. It was also used to see if their motivation increased between the pretest and posttest. The items in the IMI are listed in Appendix B.

Musical Sophistication Index. Using a shortened version of the Goldsmiths Musical Sophistication Index (Gold-MSI) [19], participants self-reported musical skills and behaviors to assess their history with musical instruments as well as a variety of items that assessed overall level of musical engagement and sophistication. The measure includes four Factors: Factor 1 is related to active engagement in musical activities; Factor 2 is related to perceptual abilities; Factor 3 is related to musical training; and General Factors is a mix of the categories. The MSI items used here are in Appendix C.

3.4. Procedure

Participants were randomly assigned to one of the three display conditions; Non-Speech, 2-Speech, or 3-Speech. The study used a between-subjects design to ensure that participants could focus on becoming familiarized with one display mapping. All sections of the study were presented via Qualtrics, and mp3 files were uploaded and integrated into the survey platform. The first task was the Listening Discrimination Task, followed by an introduction to their assigned display. Then, participants completed the pretest and filled out the first Intrinsic Motivation Inventory. Then they continued to the practice phase, which was on a separate Qualtrics survey. After practice, participants returned to the original Qualtrics survey to fill out the second motivation survey and complete the posttest. Lastly, they filled out the third and final motivation survey and the Musical Sophistication Index.

Listening Discrimination Task. Participants determined if two sound clips were the same or different.

Introduction to the display mapping. The participants were given an introduction to their assigned display. Participants clicked through example sound clips of each of the variables in their display, along with a short explanation of the parameter mapping. Participants were able to listen to the mapping examples and explanations as many times as they liked and were allowed to ask questions.

Pretest. After the participants felt comfortable with their introduction, they were directed to the pretest. The pretest evaluated the listeners' ability to comprehend the data presented within the display initially, without practice, and was used to compare to the posttest results, after practice with feedback. There were a total of 20 questions. Participants listened to a mp3 sound file embedded into the survey that combined all five variables together across the three streams. Then participants were asked to select the trend ("increase", "decrease", "constant", "increase then decrease", and "decrease then increase") of one of the variables from that sound clip. Tracks was presented in a randomized order to each participants.

Practice Phase with Feedback. The practice phase was similar to the pretest phase, but started with a short matching section to review the variable mappings. The survey also allowed participants to go back and replay the sound tracks if needed, and it provided feedback on their answers. The 20 tracks in the practice phase were similar to, but distinct from, the tracks used in the evaluation phase.

Posttest. The posttest phase occurred after practice; it followed the same procedure as the pretest, with the same 20 tracks but in a randomized order.

Motivation Checks. Participants were asked to complete the IMI scale three times: after the pretest, after practice with feedback, and after the posttest.

Musical Sophistication Index. After the participants finished the posttest and the last motivation scale, they completed the abbreviated Goldsmiths MSI.

3.5. Hypotheses

H1. The first hypothesis was (a) that there would be a difference in performance before and after the practice phase, and (b) that participants in the mixed auditory displays would perform differently from participants in the non-speech display.

H2. The second hypothesis was that individual differences such as musical experience and motivation would predict overall listeners' performance on the initial task, and would predict the amount of improvement after practice.

4. RESULTS

There were initially 102 participants in the study. Data from five were removed as statistical outliers in the pretest and posttest score; this left 97 participants for analysis. The data were analyzed with respect to the two primary hypotheses using a split-plot Analysis of Variance (ANOVA) and hierarchical linear regressions.

Table 4: Summary of Test Scores

Evaluation	Mean	Standard Deviation
Pretest	**8.68**	**2.47**
Non-Speech	9.30	2.62
2-Speech	8.66	2.34
3-Speech	8.06	2.36
Posttest	**10.16**	**2.90**
Non-Speech	10.67	3.17
2-Speech	10.00	2.82
3-Speech	9.81	2.71

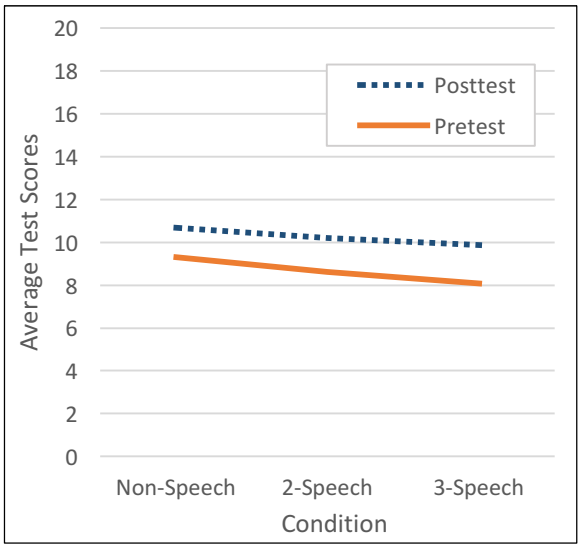

Figure 1. Average pretest and posttest score. This figure highlights the difference in average pretest scores compared to average posttest scores across the three different conditions. The range in scores is from 0-20.

4.1. Hypothesis 1

The first hypothesis was that there would be an improvement in score from pretest to posttest, and a difference in improvement between the three conditions. The results are listed in Table 4. The average pretest score across all conditions was lower than the average posttest score across all conditions. The average pretest score for Non-Speech, was higher than the average pretest score for 2-Speech, which was higher than the average pretest score for 3-Speech. The average posttest score for Non-Speech was also higher than the average posttest score for 2-Speech, which was also higher than the average posttest score for 3-Speech. Results from the split-plot ANOVA showed that there was a significant main effect for test scores $F(1,94) = 28.237$, $p < .001$, but not for condition $F(1,94) = 0.214$, $p = .807$. There was a statistically significant difference between pretest and posttest scores, but no statistically significant difference in improvement among the three conditions. These findings partially support Hypothesis 1, as there was a significant improvement in scores from pretest to posttest. This suggests that practice with feedback affected participants equally regardless of the condition, and that participants were able to improve their scores after the practice phase (Figure 1).

4.2. Hypothesis 2

The second hypothesis was that individual differences in musical experience and motivation would affect performance on the pretest and posttest scores. The results are listed in Table 5. Musical experience is the combination of the Listening Task Score and the four subsections of the Musical Sophistication Index which are Factor 1: Active Engagement, Factor 2: Perceptual Abilities, Factor 3: Musical Training, and General Factors. Motivation was measured three times throughout the study; the first one after pretest, the second after practice with feedback, and the third time after posttest.

Table 5: Summary of Individual Differences

Variable	Mean	Standard Deviation
Listening Task	20.7	6.51
Factor 1: Active Engagement	29.80	10.73
Factor 2: Perceptual Abilities	41.42	6.51
Factor 3: Musical Training	22.47	14.20
General Factors	69.23	22.85
Motivation 1	84.46	15.71
Motivation 2	86.72	16.04
Motivation 3	83.30	17.35

There were a total of fifteen step-wise linear regressions to observe the predictability of musical experience and motivation on pretest scores and on posttest scores. Additionally, pretest scores were also used to determine if they were good predictors of posttest while controlling for musical experience and motivation.

Predicting Pretest scores. For the regressions predicting pretest scores, musical experience and motivation were not significant predictors when all conditions were combined. However, in the Non-speech condition, when controlling for musical experience, motivation accounted for 37% of variance in pretest scores, $\Delta R^2 = 0.374$, $F(8,24) = 3.194$, $p = .013$.

Predicting Posttest scores. For the regressions predicting Posttest scores across conditions, musical experience and motivation were both significant predictors. Musical experience accounted for 13% of the variance in posttest scores, $R^2 = 0.131$, $F(5,91) = 2.753$, $p = .023$. There was a significant contribution of motivation when controlling for musical experience, accounting for 10.8% of variance in posttest scores, $R^2 = 0.239$, $F(8,88) = 3.460$, $p = .002$, $\Delta R^2 = 0.108$, $p = .008$. Motivation was a significant predictor of posttest scores in the 3-Speech condition, accounting for 17.3% of the variance in posttest score when controlling for musical experience, $R^2 = 0.462$, $F(8,23) = 2.465$, $p = .043$, $\Delta R^2 = 0.173$, $p = .088$. For the 2-Speech condition, both musical experience and motivation were significant predictors for posttest scores. Musical experience accounted for 24% of the variance in posttest scores, $R^2 = 0.237$, $F(5,58) = 3.602$, $p = .007$, while motivation accounted for 17% of the variance in posttest scores when controlling for musical experience, $R^2 = 0.407$, $F(8,55) = 4.712$, $p < .001$, $\Delta R^2 = 0.170$, $p = .003$.

Predicting Posttest scores with Pretest scores. The last set of regressions took the pretest score as a final predictor of posttest scores in the step-wise regression. All three predictors (musical experience, motivation, and pretest score) were significant predictors of posttest scores when all three conditions were combined. Musical experience accounted for 13% of the variance in posttest scores, $R^2 = 0.131$, $F(5,91) = 2.753$, $p = .023$, while motivation accounted for 10% of the variance in posttest score when controlling for musical experience, $\Delta R^2 = 0.108$, $F(8,88) = 3.460$, $p = .002$. Additionally, when controlling for both musical experience and motivation, pretest scores accounted for 15% of the variance in posttest scores, $\Delta R^2 = 0.152$, $F(9,87) = 6.205$, $p < .001$. All three predictors were also significant predictors of posttest scores in the two speech conditions combined (2-

Speech and 3-Speech combined). Musical experience accounted for 24% of the variance in posttest scores, R^2 = 0.237, $F(5,58) = 3.602$, $p = .007$, motivation accounted 17% of the variance in posttest score while controlling for musical experience, $\Delta R^2 = 0.170$, $F(8,55) = 4.712$, $p < .001$. Last, when controlling for musical experience and motivation, pretest scores accounted for 7% of the variance in posttest score, $\Delta R^2 = 0.071$, $F(9,54) = 5.480$, $p < .001$.

Musical Experience predicting Posttest scores. Musical experience was a combination of five variables; one listening task and four sections of the Musical Sophistication Index. Each have different standardized coefficients that can be used to determine which one is a better predictor for posttest score. In the conditions where musical experience was a significant predictor of posttest score, the coefficients of Factor 1: Active Engagement was a better predictor than the other three Factors. For instance, when data from 2-Speech and 3-Speech were combined, the first model with just musical experience as a predictor shows that Factor 1, $b = -.440$, $t(64) = -2.424$, $p = .018$ is a better predictor than Factor 2, $b = .170$, $t(64) = 1.139$, $p = .259$, Factor 3, $b = .042$, $t(64) = .211$, $p = .834$, and General Factors, $b = .443$, $t(64) = 1.544$, $p = .128$. The same trend is seen when motivation is added in as a predictor, where Factor 1 is the best predictor of posttest scores, $b = -.499$, $t(64) = -3.015$, $p = .004$, and again, when pretest scores are added as a third predictor, $b = -.474$, $t(64) = -3.018$, $p = .004$. Factor 1 is a better predictor of posttest scores than its counterparts, even when musical experience all together might not be a significant predictor.

5. DISCUSSION

This study was designed to explore listeners' ability to interpret health-related information from auditory displays before and after a practice phase, to see if practice with feedback would help improve performance and comprehension. It also investigated whether or not the display designs would have an impact on listeners' ability to improve their posttest scores after practice. In addition to looking at the effects of practice on the pretest score and posttest score, musical experience and motivation were also measured to see if any of those variables predicted scores.

Overall, practice was helpful and did improve listeners' ability to comprehend information in the auditory displays, however there was no statistically significant difference among the three conditions, Non-Speech, 2-Speech, and 3-Speech. Findings also showed that motivation and musical engagement were significant predictors of posttest scores. The remainder of this section will be split by these two main findings that correspond to each hypothesis.

The first hypothesis was that there would be a difference between pretest scores and posttest scores and that the different display designs may show different effects on that improvement between the pretest and posttest scores. This is based on the evidence that practice with feedback significantly lowers errors while performing sonification tasks [12]. It is also based on the assumption that non-speech and mixed speech auditory displays may have varying difficulty levels, each with specific design factors that can impact performance and usability overall. Findings only partially supported this hypothesis, in that there was a significant difference between pretest and posttest score but no difference among the display designs. These results indicate that regardless of the display design, the participants improved significantly between the pretest and posttest with the help of the practice with feedback. Participants generally started off scoring low during the pretest and were able to improve their score after the practice phase.

Because this task is foreign to participants, they were all starting off on the same level, where their initial performance in the tasks is generally low. However, with practice, as participants became familiar with the sounds and trends of the variables, they were all able to improve roughly the same amount across all conditions. This may also suggest that including speech into a mixed auditory display does not increase familiarity with the display compared to the non-speech display, possibly due to the fact that most participants are not exposed to mixed auditory displays, especially with multiple streams. It would be interesting to compare accuracy in monitoring change in the speech variables versus the non-speech variables to see if familiarity with speech translates to better detection of the speech variables over the non-speech variables. Additionally, workload tasks or measures of usability for each of the display designs may give better insight to how different the displays might have actually been.

The second hypothesis was that individual differences, such as musical experience and motivation, may predict how well individuals perform on the pretest and posttest. Motivation checks were a way to discard data from participants who were not motivated at all, but also because there may be a correlation between motivation scores and test scores. Findings from the hierarchical linear regressions partially supported this hypothesis, where motivation was a significant predictor of posttest scores. The effects were minimal in predicting pretest scores, most likely due to not being bored or tired yet. It serves as a good reminder that motivation plays an important role in participation and obtaining clean, representative data.

Previous research suggests that musical experience such as musical training and expertise may help listeners detect irregularities or changes in auditory streams better than those who do not have musical backgrounds [10]. Though research has not found a clear connection between musical training and stream segregation, there may still be a link that has not been found and is worth looking into [8]. In this study, musical experience included the Listening Task Score and the four sections of the Musical Sophistication Index, and was a significant predictor of posttest score. Factor 1 of the Musical Sophistication Index score is based on active engagement in music and music-related activities. Results show that Factor 1 is usually the best predictor for posttest score compared to the other factors, such as perceptual ability and musical training. This suggests that musical training and expertise is not required for monitoring auditory streams; instead, **active engagement in music** is more likely to impact listeners' ability to monitor auditory streams. In this study, those who scored high on active engagement (Factor 1) may not have had formal musical training, but could still improve significantly on the posttest, compared to someone with years of musical training. Furthermore, participants who scored high on musical training may not have scored high on motivation, while participants who scored high on active engagement may have scored higher on motivation. It would be interesting to see if active engagement correlates with motivation and interest in the study, which can lead to higher posttest scores and a larger improvement. Previous research has reached conflicting conclusions on how musical experiences impacts stream segregation and stream monitoring, but mostly because musical experience has been operationalized in so many different ways [2]. Musical training in an instrument or voice for a certain number of years may not lead to the same level of expertise or ability for each person, so using it as a

measurement may not lead to consistent results. Active engagement in music is more straightforward since it takes into account the amount of time a person spends engaging in music in a given time, while ignoring other factors such as expertise and training. These results indicate that individual differences do not matter when first introduced to an unfamiliar auditory display, but they do matter when predicting how much individuals might improve using the display with practice and feedback. This may be because the unfamiliar auditory display places everyone on the same level, but some individuals improve with practice more than others, due to individual differences.

This study scratches the surface of speech and non-speech mixed auditory display designs, and the effects of active engagement in musical activities on the usability of these displays. It demonstrates that users who actively engage in music are able to learn to use unfamiliar auditory displays better than those who do not engage in music. Continuation of this field of research can lead to better understanding of auditory display designs and training methods for future applications of these displays, such as in an anesthesiologist's workstation, a driver on a long road trip using in-vehicle interfaces, or visually impaired students using STEM education tools. Understanding how to best transform data and information into auditory streams can help reduce the dependence on visual displays and overcome information overload.

6. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

[7] 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.

[8] Schuett, J. H. (2017). Measuring the effect of display design and practice on listener accuracy for auditory displays with multiple streams (Doctoral Dissertation Proposal). Georgia Institute of Technology, Atlanta, Georgia.

[9] 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.

[10] 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.

[11] 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.

[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.

[13] Ericson, M. A., Brungart, D. S., & Simpson, B. D. (2004). Factors That Influence Intelligibility in Multitalker Speech Displays. *The International Journal Of Aviation Psychology, 14*(3), 313-334.

[14] Simpson, B. D., Brungart, D. S., Dallman, R. C., Joffrion, J., Presnar, M. D., & Gilkey, R. H. (2005). Spatial audio as a navigation aid and altitude indicator. *Human Factors and Ergonomics Society, 49*, 1602-1606.

[15] Simpson, B. D., Brungart, D. S., Gilkey, R. H., & McKinley, R. L. (2005). Spatial audio displays for improving safety and enhancing situation awareness in general aviation environments. *New Directions for Improving Audio Effectiveness, 26*, 1-16.

[16] Li, S. W., Tang, T., Hickling, A., Yau, S., Brecknell, B., & Sanderson, P. M. (2017). Spearcons for patient monitoring: Laboratory investigation comparing earcons and spearcons. *Human Factors, 59*(5), 765-781.

[17] 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.

[18] 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.

[19] Müllensiefen, D., Gingras, B., Musil, J., & Stewart L. (2014). The Musicality of Non- Musicians: An Index for Assessing Musical Sophistication in the General Population. PLoS ONE, 9(2): e89642