Advanced Auditory Menus: Design And Evaluation of Auditory Scroll Bars

Pavani Yalla and Bruce N. Walker
Sonification Lab, Georgia Institute of Technology
654 Cherry Street
Atlanta, Georgia, USA 30332-0170
+1-404-894-8265

pyalla84@gmail.com, bruce.walker@psych.gatech.edu

ABSTRACT

Auditory menus have the potential to make devices that use visual menus accessible to a wide range of users. Visually impaired users could especially benefit from the auditory feedback received during menu navigation. However, auditory menus are a relatively new concept, and there are very few guidelines that describe how to design them. This paper details how visual menu concepts may be applied to auditory menus in order to help develop design guidelines. Specifically, this set of studies examined possible ways of designing an auditory scrollbar for an auditory menu. The following different auditory scrollbar designs were evaluated: single-tone, double-tone, alphabetical grouping, and proportional grouping. Three different evaluations were conducted to determine the best design. The first two evaluations were conducted with sighted users, and the last evaluation was conducted with visually impaired users. The results suggest that pitch polarity does not matter, and proportional grouping is the best of the auditory scrollbar designs evaluated here.

Categories and Subject Descriptors

H.5.2 1 [**Information Interfaces And Presentation (e.g., HCI)**]: User Interfaces – Auditory (non-speech) feedback, graphical user interfaces (GUI), interaction styles (e.g., commands, menus, forms, direct manipulation), user-centered design, voice I/O

H.5.1 [**Information Interfaces And Presentation (e.g., HCI)**]: Multimedia Information Systems – audio input/output

K.4.2 [**Computers and Society**]: Social Issues – assistive technologies for persons with disabilities

D.2.2 [**Software Engineering**]: Design Tools and Techniques – user interfaces

General Terms

Human Factors, Design, Performance.

Keywords

auditory menus, auditory scrollbar, non-speech sounds, universal design

1. INTRODUCTION

When compared to visual menus, there is very little research involving auditory menus in the literature. However, interest in this area is increasing with the widespread use of devices such as computers and mobile phones by users who may have difficulty with the visual interface. The most obvious users for devices with auditory menus are visually impaired users, but they can also be used by sighted users who cannot look at or cannot see the screen for some reason (e.g., multitasking). Even within the user group with visual impairments, there is a wide range of vision conditions. For instance, some people are completely blind, while others have light perception and/or low vision. The range of different vision conditions makes it necessary to take a universal design approach when designing auditory menus.

While there are many guidelines that suggest how visual menus should be designed, there are no standards for the design of auditory menus. Therefore, initially, it may be useful to study visual menu concepts and determine if it is possible to devise auditory analogies to these concepts. This approach is specifically helpful when enhancing existing visual menus with audio. Users with all levels of vision could benefit from such enhancements. Those who are completely blind could rely on just the sounds, while those who have at least some vision could benefit from the visual display as well. Yet another advantage is the fact that users who lost their vision later in life might already be familiar with visual menu concepts, making the auditory menus easier to use.

2. BACKGROUND

The most common interpretation of auditory menus is that they are "speaking menus." Most people think of screen readers and text-to-speech (TTS) engines that convey the text of a menu's title and items using a synthesized voice. Although speech has the capability to convey the actual content of menus well, non-speech sounds have the potential to convey the contextual information and feedback necessary while navigating and selecting menu options. Surprisingly, there is very little use of non-speech sounds in current products. Modern visual menus, by contrast, often include much more than just the text of the menu items. Lines, icons, shortcuts, and scrollbars are all common elements. Adding analogous non-speech sounds to enhance auditory menus should therefore be systematically investigated.

Three major non-speech techniques related to auditory menus include auditory icons [6], earcons [1], and spearcons [17]. Although individual studies have been done on each of these techniques, it has recently been shown that spearcons can be more effective than both earcons and auditory icons when used as auditory menu enhancements [17]. Below is an outline of some visual menu concepts to be considered when designing auditory menus. Most of these concepts were outlined in [18].

2.1 Content Organization

2.1.1 Depth vs. Breadth Tradeoff

The visual menu concept of a tradeoff between depth and breadth belongs mainly to the category of hierarchical menus. Hierarchical menus are very common in technological devices because they allow the logical organization and presentation of a large number of menus. A hierarchical menu is a tree-like structure made up of submenus that have dependencies among themselves [10]. The selection of a single menu item determines the submenu that will be presented next. Generally, each submenu is presented one at a time, so the entire tree structure might not be apparent. Therefore, navigating through a hierarchical menu can sometimes be challenging. As the hierarchical menu gets deeper, the user must traverse more submenus before arriving at the target selection. As the menu gets broader, the user must select from more options within each submenu. Both strategies could lead to increased time taken to reach the target selection. This tradeoff between the depth and breadth of a visual menu has been studied by researchers, and even modeled with a range of techniques [5].

The results of many of these studies support broader and less deep menus as being more efficient than deeper and less broad menus [8,12]. However, it has also been suggested that the depth and breadth of a menu depend on the screen size and the complexity of the task [4]. Users actually prefer a deeper hierarchy on a mobile phone, which has a small screen size [7].

In the domain of auditory menus, there are no standards for determining the depth and breadth of a hierarchy. Since screen size is irrelevant to an auditory menu, a broader menu might be more appropriate (there is no need to worry about visual clutter). However, due to the serial quality of auditory menus, a broad menu might result in information overload for the user. Therefore, it is necessary to conduct further experiments on pure auditory menus, as well as audio-enhanced visual menus to determine the best guidelines for depth and breadth.

2.1.2 Menu Order

The items or options in a visual menu are often ordered in some logical fashion. Alphabetical, numerical, and chronological ordering are all examples of ordering techniques that can reduce menu selection times if used appropriately [10]. Ordering menus by frequency of use is another technique that has been found to significantly reduce performance times [11], although it results in dynamic menus that also lead to poorer performance due to the lack of consistency [9].

Similar research needs to be conducted with auditory menus to determine the best ordering techniques. Ordering menus by frequency of use could potentially be very beneficial, especially since auditory menus are usually conveyed serially. In most cases the user would probably be able to make a selection after listening to just a few menu items [18]. However, experiments need to be conducted to also investigate the effect of decreased consistency.

2.2 Providing Contextual Information

Contextual information such as the size and structure of the menu, as well as the user's location within it can help the user during navigation.

2.2.1 Hierarchical Menu: Structure and Location

As mentioned before, it can be difficult to navigate through a hierarchical structure with many submenus and levels. The user might feel disoriented or lost in the hierarchical structure. Therefore, conveying the overall structure of the menu and/or the user's location within the structure can be helpful. Menu maps [12] show the user the entire structure of the menu to facilitate this. A cascading menu in a desktop computer is another example of how location and structural information of a hierarchy are conveyed.

Providing contextual information in an auditory hierarchical menu is just as important. It was found that earcons [1], a sequence of musical notes, could help provide navigational cues in hierarchical menus [2]. In essence, each submenu in the hierarchy would play a different earcon in the background to help the user identify when they have moved into a different submenu. Additionally, the earcon would be composed of a sequence of sounds, each of which is characteristic of a particular previous submenu. As the user moved through the hierarchical structure, the background earcon would add an extra sound for each extra submenu that is traversed. At any point, the earcon that is heard in the background acts as a "history" of all the previous selections that were made. Although the experiments showed promising results, there were few follow-up experiments to determine the best design guidelines for the sounds to be used. Since then, a plan of study has been suggested, recommending possible designs for the earcons [18]. Experiments on these designs need to be conducted in order to determine the best design approaches.

2.2.2 Single Menu: Size and Location

Contextual information can be important within a single menu as well. If there are only a few items in the menu, its size and the user's location are apparent because all menu items are usually visible on the same screen. However, a single menu with a large number of items may require the user to scroll in order to view all of the items. This makes it difficult for the user to anticipate how long the menu is, or even know where in the menu she is currently located.

In visual menus sometimes a scrollbar is displayed to help convey this information. Scrollbars in menus are obviously used in many applications, and on many platforms. Of particular interest to the present project is the use of scrollbars in mobile phone menus, since in that context the scrollbar is purely "display", in that it is not used to actually move the cursor up or down in the list (contrast this with the scrollbar on a desktop computer application, which is both display and control). A mobile phone contacts list (or "address book") is a good example of where a menu becomes very long (easily 50 items or more), and the scroll bar can play an important role in the interaction. The location of the "thumb" of the scrollbar conveys the user's location within the menu, while the size of the thumb conveys the size of the menu. If the thumb size is small, there are many items in the menu. If the thumb size is large, there are few items in the menu.

An auditory scrollbar could be designed analogous to the visual scrollbar. This has been done previously, but with focus on enhancing a visual scrollbar as found on a computer application [3]. While that study was focused on a continuous scrollbar (the thumb can be located anywhere along the bar), most menus utilize discrete scrollbars (the thumb can only be located at discrete, designated points).

Suggestions have been made for possible designs of an auditory scrollbar that would be suitable for a single menu in a mobile phone such as the address book [18]. The focus of the rest of this paper discusses the design and evaluation of a few of these scrollbars, as well as a few new ones.

3. DESIGN OF AN AUDITORY SCROLLBAR

Four different types of auditory scrollbars were designed: Single tone, double tone, proportional grouping, and alphabetical grouping. The first two types were designed based on the suggestions laid out by [18]. They borrow many concepts from visual scrollbars, and attempt to convey the same kinds of information to the user, including size and location in a menu. They use very straightforward mappings of data onto sound, for maximum comprehension and usability. The second two types were designed based on the results of the first evaluation.

All four designs make use of tones varying in pitch played before the TTS of each menu item as the item receives focus. The tones play the role of the scrollbar. Pitch polarity is a factor that was taken into account for all four scrollbar types.

3.1 Pitch Polarity

Pitch polarity refers to the direction of pitch change as the user scrolls through menu items. For this study, increasing polarity is considered to be an increase in pitch as the user scrolls down the menu, and a decrease in pitch as the user scrolls up the menu. Conversely, decreasing polarity occurs when the pitch decreases as the user scrolls down the menu, and increases as the user scrolls up the menu. Previous research shows that the users' mental model of a particular application could result in a preference for one or the other pitch polarities [13-15]. Therefore, all of the auditory scrollbars were designed using both polarities, and pitch polarity was one of the independent variables during evaluation.

3.2 Single Tone

The single tone scrollbar is the simplest of all four designs. For each menu item that receives focus, the user hears a tone, followed by the TTS for that item. The pitch of the tone corresponding to each menu item is distinct, and the tones for consecutive items are heard either increasing or decreasing in pitch consecutively (depending on the pitch polarity chosen). Hence, the pitch of the tone heard corresponds to the location of that particular item in the menu.

The range of pitches chosen (e.g., two octaves) remains constant across different menu sizes. As the number of items in the menu increases, the pitch gap between each consecutive tone decreases and vice versa. The reasoning behind this design decision is twofold. First, this design behaves similar to a visual scrollbar. The minimum and maximum positions for the thumb of a visual scrollbar remain constant, and as the number of items in the menu increases, the position change between each consecutive item decreases and vice versa. Therefore, just as a thumb jumping down a visual scrollbar indicates very few items in the menu, a large pitch change between consecutive items in an auditory scrollbar indicates very few items in the list. The second reason for keeping the range of pitches constant is the fact that it is scalable. Consider the alternative: a design where the range is not constant, but the pitch gap between each consecutive tone is constant. Such a design could cause large menu sizes to result in pitches too high or too low for the human ear. In addition, it would not convey the size information that a visual scrollbar conveys.

3.3 Double Tone

The double tone scrollbar is a modified version that attempts to provide the user more information. Instead of hearing one tone before the TTS of each name, the user hears two tones. The pitch of the first tone would correspond to the location (just as mentioned before), and the pitch of the second tone would correspond to the pitch of the very first or very last item (depending on whether the user is scrolling up or down). The second tone would act as a reference, so if the pitch gap between the two consecutive tones is large, the user knows that she has to scroll for a long time before reaching the top or bottom of the menu. Similarly, if the pitch gap is small, it means that there are only a few items before the end of the menu.

3.4 Proportional Grouping

The proportional grouping scrollbar is a modified version of the single tone scrollbar; the user only hears one tone before each TTS phrase. However, a total of only eight musical notes are used: C4, D4, E4, F4, G4, A5, B5, C5. These eight notes were used because they would probably be the most familiar set of notes (octave of chromatic scale). If the menu has eight items, each item is assigned a note from the scale. If there are more than eight items, grouping occurs. For example, in a menu with twenty-four items, each pitch would be heard three times (i.e., the pitch would change after every three items). As soon as the user hears a group of three consecutive tones at the same pitch, she can multiply three by eight (because there are always eight notes used in any menu) to determine that there are a total of twenty-four items. This provides a quick way to estimate the total size of the menu without having to scroll through the entire list.

If the total number of items in the list is not an even multiple of eight, the remaining items are added on to the last group of three. So a menu with twenty-six items, for example, would consist of seven groups of three, and one last group of five. The reason for adding the remainder to the end is to convey the grouping information at the beginning. The main goal of this approach is to enable the user to estimate the relative length of the menu after scrolling through just a few items. When users consecutively hear multiple notes at the same pitch, they know the list is lengthier than if they hear the pitch change more frequently.

3.5 Alphabetical Grouping

The alphabetical grouping scrollbar involves grouped tones as well, but instead of changing proportionally to the size of the menu, the pitch changes each time the user scrolls to an item that starts with the next letter of the alphabet. This design is only possible when the menu items are ordered alphabetically, as they are in the address book of a mobile phone. At most, twenty-six different pitches would be used (if the menu items span all letters of the English alphabet).

4. EVALUATION STUDY 1

The first evaluation involved an examination of pitch polarity and a comparison between the single tone and double tone scrollbars. Study 1 was conducted as an interactive think-aloud session, during which the participant was asked to interact with demos of the menus, answer some questions orally, and fill-out surveys. All of the demos were different versions of an address book in a mobile phone. The address book was a menu of fifty different randomly generated names.

This particular study involved normally sighted participants, although a later evaluation (reported in Study 3) was conducted with visually impaired users. Multiple evaluations with these different user groups were conducted to support universal design principles.

4.1 Participants

Twelve undergraduate students (5 male, 7 female; mean age 19 years) participated in this study for partial credit in psychology courses. They all reported normal or corrected-to-normal hearing and vision.

4.2 Apparatus and Equipment

A Dell computer running Windows XP presented the stimuli on a 15" monitor. Sounds were presented via Sony MDR-7506 headphones, set to a comfortable listening level. A keyboard was used as the main mode of input. Specifically, the "up" and "down" keys were used to scroll through menu options. The experimenter took notes with paper and pencil. The experiment was built using Adobe Flash 8 software with ActionScript, and ran on the computer using a Flash-enabled browser.

4.3 Stimuli

The list of fifty names was derived using a random name generator (http://www.xtra-rant.com/gennames/). Text-to-speech files were generated for all of the names using the AT&T Labs Text-To-Speech Demo program (http://www.research.att.com/~ttsweb/tts/demo.php). The files were converted into MP3 format to be used in the Flash file. A user could scroll through the address book by using the "up" and "down" arrow keys on the keyboard. A visual scrollbar was also simulated as part of the address book.

The auditory scrollbar consisted of a range of fifty different pure tones. The tones were generated using the sine wave generator available in Audacity software, and the pitches ranged from 500Hz to 3000Hz. For each name the pure tone was played immediately after the key-press, and the text-to-speech was played immediately after the pure tone.

4.3.1 Part I: Pitch Polarity

The two variables in this part were the pitch polarity (defined above) and whether the visuals were left on or off. Thus, there were a total of four conditions. The order in which the conditions were presented was varied among the participants for counterbalancing.

Table 1: The four conditions of Part I of Study 1.

	Visuals On	Visuals Off
Increasing Polarity	A	B
Decreasing	C	D

4.3.2 Part II: Single Tone vs. Double Tone

There was only one variable in this part, and therefore two conditions: single-tone and double tone. The single-tone stimulus that was shown was based on whichever pitch polarity the user had preferred in the first part of the study. The double-tone stimulus was created by adding a second tone (either the maximum or minimum pitch value, depending on the direction of scrolling) between the first tone and the text-to-speech.

Table 2: The two conditions of Part II of Study 1.

	Visuals On
Single Tone	A
Double Tone	B

4.4 Procedure

First, each participant was asked to sign a consent form and fill out a demographic questionnaire including the participant's age, gender, academic major, and number of years of musical experience/training.

4.4.1 Part I: Pitch Polarity

After the interviewer explained the purpose of the research, the participant was asked to put on headphones and use the up/down keys on the keyboard to scroll through the menu on the screen. The participant was not given a specific task, but was asked to scroll through the entire menu at least once. After interacting with the first menu (Option 1), the participant was asked to interact with another menu that was identical to the first, except that the pitch polarity of the scrollbar was reversed (Option 2). When the participant had finished interacting with both menus, the interviewer asked if the participant could identify the difference between the menus. Then the participant was asked to fill out a survey and was asked about which option (i.e., which polarity) she preferred. This was repeated again with the visuals turned on/off (depending on which condition was shown to them first).

4.4.2 Part II: Single-tone vs. Double-tone.

The interviewer then presented the double-tone scrollbar to the participant, and asked her to interact with it as she had done with the other options. Subsequently, the participant was asked whether she could tell the difference between this new (double-tone) option and the previous (single-tone) ones. The participant was asked to fill out another survey and was asked whether she preferred the new option or the old ones.

The participant was also asked to fill out an additional survey to give the researchers more feedback about the concept of enhancing visual menus with audio. Finally, the participant was asked to give comments/suggestions about the specific designs presented here. The interviewer recorded any comments that the participant made throughout the session. Additional details about the surveys and stimuli are available from the authors.

4.5 Results

4.5.1 Part I: Pitch Polarity

All of the participants were able to identify the difference between the increasing and decreasing pitch polarity conditions. Of the twelve participants, half preferred the increasing pitch polarity (low to high), while the other half preferred the decreasing pitch polarity (high to low). More importantly, five of the six who preferred increasing polarity were shown the increasing option first, and five of the six who preferred decreasing polarity were shown the decreasing option first. Therefore, in general the participants preferred whichever option they heard first. Only one of the participants reported a difference in preference between the visuals-on and visuals-off conditions.

Many of the participants commented that the high-pitched tones were too high. They suggested that the entire pitch range be shifted lower. As explained, a wide (and even overly wide) range of pitches is a likely result of using the one item-to-one pitch menu design. Thus, even shifting the whole range down might not be a completely satisfactory solution to this concern. Two participants reported that they preferred the decreasing polarity because the increasing polarity was "distracting" and "annoying" as they scrolled down, and the pitch became increasingly higher.

4.5.2 Part II: Single-tone vs. Double-tone

In general, most of the participants could not completely identify the difference between the single-tone and double-tone scrollbars. Most could identify that two tones were heard with each movement, but they did not understand that the second tone was a reference tone of a minimum or maximum. Ten of the twelve participants preferred the single-tone option to the double-tone option. Most of them thought that the second tone was "unnecessary" and "distracting." However, some of them further commented that their opinion is probably different from that of a blind user's. They noted that the double-tone scrollbar could potentially benefit a blind user. That is, participants seemed to recognize that a more complex or more sophisticated display might be useful for some users, but they did not see the value in it for their own usage, at least not at this level of familiarity with the auditory menu and auditory scrollbar.

5. STUDY 2

The second evaluation was a comparison between the proportional grouping and alphabetical grouping scrollbars. The designs of these scrollbars were actually motivated by the results of Study 1. Since most participants had preferred the single tone to the double tone, in Study 2 both of these scrollbars played only one tone before each TTS name. However, the mapping of pitches to items was different, using the proportional grouping or the alphabetical grouping (as described in a previous section).

The procedure was similar to that of Study 1, but in addition to subjective feedback from the participants, objective data was also collected.

5.1 Participants

Ten undergraduate students (6 male, 4 female; mean age 20 years) participated in this experiment for partial credit in psychology courses. They all reported normal or corrected-to-normal hearing and vision. None had participated in the first evaluation.

5.2 Apparatus and Equipment

The apparatus was the same desktop computer setup as was used in Study 1.

5.3 Stimuli

The address book was constructed in the same way as it was in Study 1. However, three different menu sizes were created for each scrollbar type: 8, 24, and 48 names.

Since the visuals on/off condition did not seem to influence participants in the Study 1, it was not included as an independent variable in this evaluation. The visuals were kept off for all conditions. However, a new dependent variable was introduced: accuracy of the participant's estimate of the menu size.

5.4 Procedure

After following the consent procedures and obtaining demographic information, the interviewer presented one of the scrollbar types to the participant, and explained the logic behind its design. Then the participant was asked to interact with the menu for ten seconds and estimate how many names were in it. This was repeated for all three sizes of each scrollbar type. Half of the participants experienced the proportional grouping scrollbar first, and the other half experienced the alphabetical grouping scrollbar first (for counterbalancing). The different menu sizes were presented in a different order for each scrollbar type. Along with this objective data, participants were asked for their subjective feedback about which scrollbar they preferred.

5.5 Results

Both the objective and subjective data suggest that proportional grouping is better than alphabetical grouping. First, all ten participants were able to estimate the menu sizes more accurately for proportional grouping. More specifically, response error was defined as the absolute percentage error of the menu size estimates. That is, the absolute difference between the estimate and the actual number of items, divided by the actual number of items, all multiplied by 100. For each participant, mean error was computed for each condition.

Overall average error (collapsing across participants) was 10.1% (*SD* = 15.73) for the proportional grouping condition, and 47.4% (*SD* = 28.37) for the alphabetical grouping condition.

This difference is statistically reliable, based on a paired-samples t-test, $t(9) = 4.23$, $p = .001$.

From the subjective data, all but one participant preferred the proportional grouping to alphabetical grouping. That one person who preferred the alphabetical grouping did so because he reported being more familiar with it. Some also observed that alphabetical grouping could be misleading because of the possibility of a very irregular distribution of alphabetical names in a list (e.g., 10 names starting with 'A', no 'B's, 1 'C', 100 'D's, etc.). They also suggested that alphabetical grouping might be better for conveying location information (where in the menu the user is currently located), while proportional grouping would be better for conveying size information (how many menu items). However, the information provided by alphabetical grouping is also available by listening to the beginning of the TTS name. Therefore, some participants thought it might be redundant.

6. STUDY 3

Study 3 was a combination of the first two evaluations, but conducted with visually impaired participants. The alphabetical grouping scrollbar was left out of this study, because it was not received well in Study 2. That left single tone, double tone, and proportional grouping designs (all with both polarities possible). Also, the evaluation was conducted using a mobile phone instead of the desktop simulation that was used in the previous studies.

6.1 Participants

Eight visually impaired adults participated in this study (4 male, 4 female, mean age 50 years). Their visual impairments ranged from low vision to completely blind. All were recruited through the Atlanta Center for the Visually Impaired and received $10 compensation for their participation.

6.2 Apparatus and Equipment

The menus and sounds were presented, and data were collected, using a Java program that simulated the list of names that would be found in a contact list on a mobile phone. Settings allowed the experimenter to change the type of menu, number of menu items, and type of auditory enhancement(s). The software was installed onto a Nokia N95 mobile phone, and users interacted directly with the device. Data logs were used to record the interactions.

6.3 Stimuli

The single tone, double tone, and proportional grouping scrollbars were designed as described in the previous section. Once again, three different sizes were created for each scrollbar type: 8, 20, and 48 names. The pitch of the scrollbar tones was lowered overall, based on the feedback received from participants in the first evaluation. The new range of tones was from 130.8 Hz (MIDI note #48) to 1975.5 Hz (MIDI note #95).

Since the participants were visually impaired, there was no need for the visuals off condition. That is, the level of their own vision dictated how much of the visuals they would see or not see (which would be the case in a real life application).

6.4 Procedure

The consent procedure differed slightly in that a sighted witness was present to sign the consent form along with the participant. Also, in addition to the demographic information collected from participants in the previous evaluations, some extra demographic information was collected: details about vision, age of onset of visual impairment, and extent of mobile phone usage.

First, the interviewer presented the single tone scrollbar with both pitch polarities. The participant was asked for their preferred pitch polarity. Then the participant was asked to interact with the menu for ten seconds and estimate how many names were in it. This was repeated for all three sizes of each scrollbar type. The different menu sizes were presented in a different order for each scrollbar type. Along with this objective data, the participants were asked for their subjective feedback about which scrollbar they preferred. The interviewer then assisted the participant in filling out the same surveys that were used in the previous evaluations (by reading the questions and noting the answers). Finally, the participant was asked open-ended questions about how she would like to improve the scrollbar designs or anything else on their own mobile phone.

6.5 Results

The results for preferred pitch polarity were quite evenly split: four preferred decreasing polarity (in which the pitch descends as the user scrolls down the list); three preferred increasing polarity; and one reported no preference.

Both the objective and subjective data collected suggest that proportional grouping is better than both single tone and double tone. Error was defined as in Study 2. Average error (collapsing across participants) was lowest for the proportional grouping condition: mean error was 5.7% ($SD = 2.0$) compared to 51.9% ($SD = 66.0$) for the single tone condition, and 75.3% ($SD = 68.6$) for the double tone. In fact, all eight participants were able to estimate the menu sizes more accurately for proportional grouping. A repeated measures ANOVA on error, with condition as the within-subjects grouping variable, revealed a significant effect of condition, $F(2, 14) = 4.00$, $p = .086$. Pairwise comparisons indicated that the proportional condition led to significantly lower error than both the single tone and double condition, $p < .05$, whereas there was no significant difference between the single and double tone conditions, $p > .05$.

However, the subjective data were more split: five preferred proportional grouping, three preferred single tone. Most of the participants found the double tone scrollbar to be confusing (and this is also supported by the error data). The participants who performed better with the proportional grouping scrollbar, but still preferred the single tone, generally did not like doing the mental calculation to determine menu size for the proportional grouping.

7. DISCUSSION AND CONCLUSIONS

As auditory menus become more and more common, making them as sophisticated as visual menus will require careful research and design efforts. In the work presented here, we have used an analysis of modern visual menus for inspiration in terms of what information needs to be included in advanced auditory menus, and evaluated various designs for preference and for performance.

Overall, in the studies presented here the proportional grouping scrollbar resulted in the best performance and received the most positive subjective feedback. It seems to be a very effective and acceptable auditory menu enhancement, and should be considered when deploying these kinds of systems.

A basic single tone scrollbar was preferred over double tones, though the higher information content in the double tone may be useful for some listeners, possibly with some training. That remains an open question for future research. Study 2 showed that the alphabetical grouping scrollbar was not very effective, and was not preferred, so it did not receive further study.

In these studies, pitch polarity did not seem to matter much when designing an auditory scrollbar for names. Participants seemed to prefer whichever polarity they heard first. Since the polarity variable was counterbalanced, this led to about half of the participants preferring each polarity. It may be that different kinds of menu content might elicit different polarity preferences, as has been found for the polarities used with other auditory displays [13-15]. The address book content used here (i.e., names in alphabetical order) does not have a particularly strong spatial or size ordering, but other kinds of menu data might (e.g., perhaps "months" or "sizes of cities"). Care must be taken to pilot test all such sound mappings [see 13-15], especially including listeners from the real user population [e.g., visually impaired, see 16].

Participants who had musical training seemed to grasp the meaning behind the changing pitches quicker than those without any musical training. However, those without musical training still performed well, given sufficient explanation of the theory behind the changing pitches. One notable suggestion received from a visually impaired participant who had no musical training was to use a total of 10 notes instead of 8 notes (musical octave) for the proportional grouping scrollbar. He said that it would be much faster to multiply by 10 rather than by 8 when predicting the size. However, this is in contrast to a different participant with some musical training who suggested that the 8-note "standard" octave be used, rather than 10 notes.

Based on the answers to the survey questions, both sighted and visually impaired users found the sounds to be helpful and informative. Naturally, the visually impaired users seemed the most enthusiastic about the auditory scrollbar. The interviewer reported that of the visually impaired users, those who were not as enthusiastic about the scrollbar were those who had lost their vision at early childhood or birth. They seemed to have found ways to get by, which they were very used to by now. For example, one participant who has been blind from birth stated that she never even used the address book function in her cell phone. She just memorized the numbers that she needed to call and dialed them manually. She had been doing this for all of her life, and it had so far proven satisfactory. In contrast, those who had lost their vision more recently were very excited to start using the new enhancements. They were eager to be able to receive the same information that they used to receive while navigating menus when they had normal vision. Also, those participants who were already familiar with the notion of a visual scrollbar (before they lost their vision) performed better and gave more positive feedback than those who were blind for most of their life and had never seen a visual scrollbar.

These observations provide important implications for the design of auditory scrollbars and auditory menus in general. Using analogies to visual menus seems to be an appropriate way of producing a universal design. It makes it easier for users with normal or low vision to use the auditory version. Even users who become completely blind may be able draw from previous experience with the visual counterpart. In the end, a universally designed, auditory enhanced menu system should lead to increased access to, and effective use of, today's increasingly sophisticated everyday technology, by all users.

8. ACKNOWLEDGMENTS

We thank the Atlanta Center for the Visually Impaired for helping recruit visually impaired participants and allowing us to conduct evaluations at their facility. Portions of this research were supported by Department of Education (NIDRR) funding of the WirelessRERC. We thank Nokia for the donation of mobile phones.

9. REFERENCES

[1] Blattner, M., Sumikawa, D. & Greenberg, R. (1989). Earcons and icons: Their structure and common design principles. *Human Computer Interaction* 4(1): 11-44.

[2] Brewster, S. Raty, V., & Kortekangas, A. (1996). *Earcons as a Method of Providing Navigational Cues in a Menu Hierarchy.* Paper presented at the HCI'96 Conference, Imperial College, London, UK.

[3] Brewster, S. A., P. C. Wright, et al. (1994). *The design and evaluation of an auditory enhanced scrollbar.* Paper presented at the CHI'94 Conference, Boston, MA, USA.

[4] Chae, M. & J. Kim (2004). Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. *Behavior and Information Technology* 23(3): 165-181.

[5] Eberts, R. (1997). Cognitive modeling. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2e). New York: Wiley. (pp. 1328-1374).

[6] Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*. 2, 167 - 177.

[7] Geven, A., R. Sefelin, et al. (2006). *Depth and Breadth away from the Desktop – the Optimal Information Hierarchy for Mobile Use*. Paper presented at the MobileHCI'06 Conference, Helsinki, Finland.

[8] Kiger, J.I. (1984). The depth/breadth trade-off in the design of menu-driven user interfaces. *International Journal of Man-Machine Studies*, 20, 201-213.

[9] Mitchell, J., & Shneiderman, B. (1989). Dynamic versus static menus: An exploratory comparison. *SIGCHI Bulletin* 20(4): 33-37.

[10] Norman, K. L. (1991). *The psychology of menu selection: Designing cognitive control at the human/computer interface*. Norwood, N.J.: Ablex Publishing Corporation.

[11] Sears, A., & Shneiderman, B. (1994). Split Menus: Effectively Using Selection Frequency to Organize Menus. *ACM Transactions on Computer-Human Interaction (TOCHI)* 1(1): 27-51.

[12] Shneiderman, B. (1986). Designing menu selection systems. *Journal of the American Society for Information Science* 37(2): 57-70.

[13] Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied, 8, 211-221*.

[14] Walker, B. N. (2007). Consistency of Magnitude Estimations With Conceptual Data Dimensions Used for Sonification. *Applied Cognitive Psychology, 21,* 579-599. DOI: 10.1002/acp.1291.

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

[16] 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.

[17] Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speech-based Earcons Improve Navigation Performance in Auditory Menus. Proceedings of the International *Conference on Auditory Display (ICAD 2006)*, London, England (20-24 June). pp. 63-68.

[18] Yalla, P., & Walker, B. N. (2007). *Advanced Auditory Menus*. (No. GIT-GVU-07-12.): Georgia Institute of Technology GVU Center.