Speech and Non-Speech Audio: Navigational Information and Cognitive Load

Anna Rouben

Research In Motion
1180 NW Maple Street, Suite 330
Issaquah, WA 98027 USA
arouben@rim.com

Loren Terveen

University of Minnesota
Department of Computer Science and Engineering
Minneapolis, MN 55455, USA
terveen@cs.umn.edu

ABSTRACT

Cell phones and other mobile devices let people receive information anywhere, anytime. Navigation information – directions and distance to a destination, interesting nearby locations, etc. – is especially promising. However, there are challenges to delivering information on a cell phone, particularly with a GUI. GUIs aren't ideal when a person's visual attention is elsewhere, e.g., scanning for landmarks, assessing safety, etc. And they don't work at all for blind people, who particularly need navigation assistance.

Our work responds to this challenge. We investigate the use of two non-visual techniques for delivering navigation information, speech and sonification [[3], . We conducted an experiment to compare user performance with and preference for the two techniques, in both single task (navigate to a target) and dual task (navigate to a target and respond to an auditory stimulus) conditions. Users performed better with and preferred sonification in both conditions. We discuss the implications of these results for the design of navigation aids.

[Keywords: Navigation, sonification, speech, cognitive load, secondary task]

1. INTRODUCTION

Mobile devices have raised the promise of the right information being there for you whenever you need it, wherever you are. Applications have been built that let people find nearby friends [2], get information about places of interest [2], and remind themselves of tasks they need to do [14], among other uses. One particularly useful type of information is navigation assistance. For example, if you're driving in an unfamiliar area, a timely prompt can help you take the right exit.

Navigation aids are even more useful for blind people. Full participation in modern society requires independent mobility. It's so simple that most people take it for granted, but finding one's way across campus to a class or locating one's doctor's office in a large medical center be difficult or impossible for people who are blind or visually disabled. Existing aids such as the white cane and guide dog are of great help, particularly in avoiding obstacles along the way. However, they don't solve other key problems faced by visually disabled people, including route planning, learning spatial layouts, discovering landmarks, and delivering information necessary to navigate to a destination.

Graphical users interfaces (GUIs) are not a good choice for delivering navigation information on mobile devices. First, blind people cannot use them. Second, sighted people typically have their visual attention elsewhere, whether while driving (on the road) or walking (e.g., scanning the environment for landmarks, familiar faces, or potential dangers). Thus, using a GUI would disrupt their attention.

This reminds us that navigation doesn't happen in isolation. While traveling to a destination, one might be driving and following traffic signs, or walking and talking on a cell phone. These activities all take cognitive resources. Concurrent secondary tasks can create potential hazards. For instance, using cell phones while driving increases time to notice and react to road signals and dangers. Therefore, it is important to find an information delivery technique that reduces distraction and imposes minimal cognitive demands.

There are three types of non-visual information delivery techniques to consider: speech, sonification, and haptics (tactile output). In the research reported here, we implement and experimentally compare two techniques – speech and sonification. We do this not just when navigation is the only task, but also when users have to attend to another auditory task. We are particularly interested in secondary tasks that involve listening to speech, since navigation often is done while listening to speech: people walk while talking to friends, and drive while talking to passengers or listening to the radio.

Many factors influence people's need for navigation information: familiarity with an area, mode of transport (walking, driving, bicycling), type of area (inside or outside, city streets vs. college campus), and degree of visual ability (from fully sighted to completely blind). In the current research, we have abstracted away from these factors, concentrating instead on fundamental issues of how people perceive audio information.

To evaluate the use of speech and sonification for delivering navigation information, we designed a simple 2D virtual space (see Figure 1). The space contains a single target destination. Users move in the space with a mouse or other pointing device. Speech or sonification instructions direct users toward the target. Obviously, this virtual space is much simpler than the physical world; e.g., real space is 3D, contains obstacles, typically there are environmental sounds, etc. Nevertheless, it is sufficient to let us make basic comparisons of the effectiveness of speech and sonification for conveying navigation information.

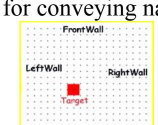

Figure 1. Virtual space with a single target object

We evaluated the performance of speech and sonification systems in a controlled experiment. Twenty blindfolded1 subjects navigated to the target from several different starting points. They then repeated the process while doing a secondary auditory comprehension task, letting us investigate the extent to which speech and sonification impose cognitive demands and distractions.

The remainder of the paper is organized as follows. We begin with a discussion of related work in speech and sonification technology and navigation aids, and then introduce our experimental platform. The heart of the paper describes our experiment, research questions and methods, followed by our results. We conclude with a brief summary and a discussion of future work.

2. BACKGROUND & RELATED WORK

2.1. Sonification output

To create an effective sonification it is important to choose a correct polarity and mapping of data dimensions to the audio dimensions [26]. Positive polarity means that changes in both sonification parameters and data occur in the same direction (higher pitch → higher temperature). An increasing pitch identifies increasing temperature and vice versa. The change in different directions (higher → lower) creates negative polarity. Additional considerations must be taken into account when designing sonifications for visually disabled people. Studies have shown that blind and sighted people largely agree in their perception of polarities, but there are exceptions. The study by Walker and Lane [25] showed that sighted and non-sighted participants agreed on the polarities representing temperature, velocity, and pressure, but disagreed on the polarity for money. Since this is quite different information than what a navigation aid must convey, it remains an open issue whether sighted and blind users will agree on polarities for navigational sonifications.

Successful sonifications have been created for monitoring (in hospital environments), data analysis, and exploration tasks [9]. Sonification for a medical workstation [5] is an example of an effective audio display; medical students identified an emergency situation more quickly than visual and audio+visual displays. A pilot study of the Marketbuzz system [22], representing real-time financial data used by financial traders, demonstrated that sonification increased accuracy in monitoring the data change. Another successful non-speech audio system was created for representing geographical information **Error! Reference source not found.**; users were able to recognize geographical data distribution patterns.

Currently there are very few guidelines for designing effective sonification. Even though sonification designs are created for a variety of data, there has been little research on the use of sonification to support navigation in geographical space. Thus, we complement previous work by focusing on this problem.

1 Blindfolding simulates blindness, a common method in studies of assistive technologies for blind people. Obviously, it would have been trivial for sighted people to get to the target. This study did not attempt to test how well sighted people could use speech or sonification in situations where their visual attention was occupied elsewhere. As discussed below, that is a matter for future work.

2.2. Speech output

Speech has several advantages for navigation systems. It is very good at expressing precise information, e.g., times (“1 hour and 7 minutes”), distances (“2.5 kilometers”), or place names (“Walter Library”). People already understand speech and do not need to learn new words and expressions to follow spoken instructions in a language they know. Speech is very expressive, so it can represent a wide variety of information (“The gym closes at 3 pm every Friday”).

Speech has shortcomings, too. Understanding speech requires active listening and interpretation, and thus can draw attention away from any other concurrent tasks. For example, it is hard to talk on the phone while actively listening to the news on the radio. It is hard to communicate dynamically changing information. Suppose you are walking quickly and would like to hear whether each step takes you closer to a target location. You might take several steps while still listening to a message telling you whether your third step in the past was in the right direction. Making sense of speech messages for navigation may require use of short-term memory. For example, if a sequence of navigational messages includes items such as “125 meters to the target” and “134 meters to the target”, you need to remember previous items and compare several items to tell whether you’re going in the right direction. The longer the messages are, the harder they are to remember [24]. Finally, navigation messages typically must be synthesized. Pronunciation and speech quality also affect the amount of effort needed to process speech messages [[20], [11]]. While synthesized speech has improved dramatically, it still doesn’t sound quite natural; therefore, it can be more difficult to process than natural speech.

2.3. Comparison of speech and sonification

Because of these shortcomings, we wanted to evaluate the use of sonification to deliver basic navigational information. Units of sonification are shorter than most words in a language. A unit can be a single musical tone, i.e. a tone coming from the right speaker means that the target is on the right. A sequence of sonification has many properties – volume, tempo, pitch, balance, and reverberation – that can be manipulated to represent information. For instance, temperature could be represented by pitch, with higher pitches meaning higher temperatures and lower pitches lower temperatures. Factors such as system properties, environment, and the type of information that must to be communicated usually determine the choice of audio type [21]. Speech is better (or the only choice) for symbolic information and precise data, such as location and event names, times, or absolute distances. Sonification’s shorter units make it well suited for conveying rapidly changing data such as relative distance and orientation.

2.4. Navigation assistance systems

Significant work has been done on developing navigation aids for blind users. As mentioned before, the results of this work also are useful for sighted users when their visual attention is occupied with other activities. Most state of the art navigation aids convey information to the user via speech or electronic Braille displays. Given our focus, we concentrate on speech output systems. We also discuss several systems that use sonification and haptic output.

Navigation System for the Blind by Loomis [12] employs spatialized synthesized speech messages. A user hears a speech message as if it was coming from a certain location. Directional spatialization was accomplished with virtual displays that utilize head-related transfer function (HRTF) **Error! Reference source not found.** also known as anatomical transfer function (ATF) to produce the appropriate signal. HRTFs simulate more realistic 3D sound with headphones than with speakers. One disadvantage of using headphones in the navigation system is that they block environmental cues, which are also helpful for both sighted and non-sighted people [23].

The Robotic Guide [10] consists of various sensing and computational equipments placed on a platform with wheels. Users hold a leash attached to the platform and follow the robot. The movement of the Robotic Guide tightens or releases the leash giving tactile feedback to the user. Users interact with the Robotic Guide through speech and a wearable keyboard. The Robotic Guide gives users auditory feedback in the form of speech and auditory icons [6]. Auditory icons convey information about an object or an event using sound effects closely related to an actual object or event, i.e. a water fountain can be represented using the sound of running water.

Our research compares speech and sonification for delivering navigation information. Similar work was done by Loomis [13], who compared virtual tones to speech. The virtual tones were spatialized to make the sound appear to come from the location of the target. The system played an on-course tone, which changed to a "whooping" sound when the user went in the wrong direction. Periodically, the tones are accompanied with speech messages conveying the distance. Experimental evaluation found that speech did not perform significantly better than virtual tones. The two techniques were evaluated for a single navigation task. Our research extends that of Loomis by exploring different sonification techniques and evaluating speech vs. sonification when a secondary task is introduced.

A few systems also utilize tactile feedback. One major advantage of tactile feedback over audio is that it does not block ambient sounds, which can be very important during navigation. For example the noise of cars informs a person about an intersection, or the sound of footsteps indicates an approaching person.

The CyARM system [17] uses tactile output to convey information about distance. Haptic feedback [15] is delivered with a motor-wire attached to the user's belt. When the distance between the user and an object decreases, the wire shortens and vice-versa. On the one hand, this method seems intuitive because it has the natural mapping of a hand extending to reach an object. On the other hand, it is not feasible in situations when arms are busy with other tasks (holding a phone, carrying a bag, using a steering wheel) because the motor-wire constrains arm movement. People with visual disabilities often like to use a white cane and/or a guide dog along with any technological aid. Restricting their arms' movement might require abandoning existing aids. Audio feedback does not have this limitation so is worth exploring.

The Haptic Eyes system [16] assists users in navigation by using different types of vibrations. Haptic Eyes collects visual data and converts corresponding images into haptic cues. These cues are delivered to the users by the mobile phone vibrators attached to the shoulders. Vibrating signals represent specific types of information, i.e. stop-danger (strong signal on both shoulders), or turn left (strong signal on the left shoulder, weak signal on the right shoulder). Even though this approach seems promising, we chose to primarily focus on audio feedback. The dimensions of audio feedback (vocabulary, sound properties) are richer than dimensions of vibration and thus might be more suitable to provide sufficient navigational information.

3. EXPERIMENTAL PLATFORM FOR EVALUATING THE DELIVERY OF NAVIGATION INFORMATION

For our purposes, the task of the navigation assistance system is to convey the information necessary to reach a target. (In a complete system, other types of information, e.g., concerning nearby obstacles and hazards, also must be conveyed. We do not address this in the current research). We concentrate on two basic types of information: distance to the target and orientation to the target.

In our 2D virtual space, the user's current position is represented by the cursor position. Distance to the target is represented in pixels. This was suitable for our purposes because we needed to convey to users whether they were moving closer to or further from a target. Thus, absolute distance and the distance unit really didn't matter. Orientation to the target has four possible values: up, down, left, and right. We also want to notify users when they run into one of the "walls" that define the edges of the virtual space.

We next discuss how distance and orientation information is conveyed in our two different techniques, speech and sonification.

3.1. Speech

As users move through the virtual space, a synthesized male voice continually tells them their distance and orientation from the target. Messages articulate orientation + distance, e.g., "left 23" or "right 15". When the user is directly above or below the target, no orientation information is provided, e.g., a message simply would be "42" or "87". Users figure out whether the target is up or down just by moving (the mouse) up or down in the space, since this makes the distance reported increase or decrease. We did not include "up" or "down" orientation in the speech output to make it equivalent to the sonification output (see discussion below).

3.2. Sonification.

A sequence of tones is generated continuously as a user moves through the virtual space. The properties of the tones (pitch and tempo) are dynamically modified to represent the user's distance to the target object. Both increase when users are moving closer to the target and decrease as they move away – higher pitch and faster tempo → closer, lower pitch and slower tempo → further.

To communicate the distance we chose pitch and tempo because they map well to human steps. Quicker steps towards the target indicate decreasing distance and time to reach the target. When you increase the speed of the steps, the distance to the target decreases more quickly. In this case, tempo and pitch polarity is negative; increasing tempo and pitch correspond to decreasing distance and time needed to reach the target. The study will help to verify whether pitch and tempo do indeed map naturally to the distance between the user and the target.

Orientation to the target is represented with a simple audio spatialization (panning): when the target is left of a user, the sound stream comes only from the right speaker/headphone, and when the target is to the right, sound comes only from the left. When the target is directly above or below the user, sound comes from both the left and right. The source of the sonification sound (left or right speaker) maps naturally to the direction of the sound coming from the target object. The study will attempt to show whether a simple spatialization (left, right, or both speakers) is effective to communicate the orientation of the target.

This simple technique doesn't distinguish whether the user is above or below the target; however, as we mentioned, users can move the mouse to determine which is the case. Follow up work can create sonifications to convey orientations including "forward" and "back" as well as "up" and "down," which are relevant when stairs and elevators are present. Clearly, additional audio properties could be used to allow a richer distinguishing of orientation; however, we found our simple technique both acceptable to users and sufficient to investigate our research questions.

Our designs of speech and sonification feedback, even though limited, convey equivalent information: relative distance and orientation. We believe that this simple design allows us to evaluate the relative utility of speech and sonification output and identify the advantages of both audio types.

3.3. Walls

Both systems notify users when they "run into" one of the walls of the 2D space with one of the following spoken messages: "left wall", "right wall", "front wall", and "back wall".

3.4. Implementation

We used the Java sound API 1.4.2 [19] and FreeTTS 1.2 [1]speech synthesizer to generate speech and manipulate the sound properties on the test interface, while the virtual space and user interaction was implemented with the Java Swing GUI toolkit [8].

4. RESEARCH QUESTIONS

We organize our work around the following questions.

RQ1. Is sonification more effective than speech at delivering basic navigation information (distance and orientation to a target destination)?

RQ2. Is sonification more effective than speech at delivering basic navigation information in the presence of a secondary (speech comprehension) task?

RQ3. For both single and dual task conditions, what are subjects' subjective preferences for speech and sonification?

5. THE EXPERIMENT

5.1. Subjects

Subjects included 20 university students (10 male and 10 female). No subjects reported any hearing problems. None of the subjects were visually disabled, and they were blindfolded during the experiment.

We are aware that the perceptions of sighted people can be different from the perceptions of people who are blind. People with different forms of visual disabilities also need different information to navigate effectively [4]. Nevertheless, we speculate that it is likely that feedback (speech or sonification) that is less cognitively demanding for sighted individuals is also less cognitively demanding for blind individuals. Of course, to verify this assumption, future research studies, which include blind participants, are needed.

5.2. Design

The design was within-subjects and each subject carried out 20 trials. Each trial required the subject to navigate from a start point to a target. Trials were grouped as shown in Table 1.

Single Task	{Speech, Sonif}	$ST_1, ST_2, ST_3, ST_4, ST_5$
	{Speech, Sonif}	$ST_1, ST_2, ST_3, ST_4, ST_5$
Dual Task	{Speech, Sonif}	$ST_1, ST_2, ST_3, ST_4, ST_5$
	{Speech, Sonif}	$ST_1, ST_2, ST_3, ST_4, ST_5$

Table 1: Experimental tasks {ST1 – starting point 1, ... , ST5 – starting point 5}

First, we examined the performance of speech and sonification when the participant had only the single task of navigating to a target destination from different starting points. Second, we evaluated speech and sonification when the participant had to do two tasks simultaneously: navigate to the target and respond to a spoken sequence of letters (details to follow). Within the single task and dual task conditions, speech and sonification were counterbalanced: half the subjects heard speech first, and half heard the sonification first. The counterbalancing was independent for the single task and dual task. Within each of the main four conditions (Single-Task / Speech, Single-Task/Sonif, Dual-Task/Speech, Dual-Task/Sonif), subjects did 5 navigation trials. In each trial, subjects were positioned at a start position in the space and had to navigate to a target. The target was the same in all trials, and there were 5 different starting points ST1, ST2, ST3, ST4, ST5 (see *Figure* 2). Each starting point was the same distance from the target, and the sequence in which subjects commenced from the starting points was the same across subjects and conditions.

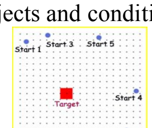

Figure 2. Virtual space configurations used for the experiment: one target and five different start points

5.3. Secondary Task.

For the secondary task, we selected a simple version of the N-Back task [18] that introduced additional cognitive load. Subjects had to listen to a spoken sequence of letters and give a verbal response when the sequence "k, k" occurred. For example, a sequence "c, m, n, k, k, b, k, p, k, k" should trigger two responses. All subjects heard the same sequence of letters.

5.4. Performance Measures

For both the single task and dual task, time to reach the target was measured. (All subjects reached the target on all trials). For the dual task, we also counted the number of correct and missed responses.

5.5. Equipment

Two laptop computers were used during the study. The first computer contained the software that generated the 2D space and produced the navigation audio output. Subjects heard navigation sounds via stereo headphones worn over both ears. Other types of audio output devices that users might prefer are also available: small clip-on shoulder- or collar-mounted speakers, headphones worn near the ear, or tubelike headphones worn in the ears [7]. Headphones and speakers that are not worn inside or over the ears block ambient sound less and therefore are potentially better for navigation.

The second computer generated the spoken sequence of letters for the secondary task. This audio was articulated with a female voice and came from speakers. This was done to emulate what we consider a reasonable usage scenario for a navigational assistance: a user is receiving navigational information through headphones, while potentially needing to attend to ambient sound or speech from the environment.

5.6. Survey

After completing the Single Task condition, and again after completing the Dual Task condition, subjects completed brief surveys. They were asked to judge factors such as ease of use of the two output techniques, audio quality, difficulty of doing two tasks simultaneously, and subjective factors such as fun. Most survey questions were evaluated based on the Likert scale (1-Strongly Disagree to 5-Strongly Agree).

6. RESULTS

6.1. Single Task

Subjects navigated to the target significantly faster with sonification than with speech. For sonification, the mean time was 20.6 seconds, and for speech, the mean time was 26.9 seconds (t-test; n=100, df=198, t=2.36, p=0.02). Figure 3 shows the mean times for subjects to reach the target for the five different start configurations.

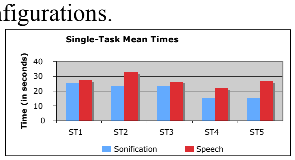

Figure 3. Mean time to target during the single-task condition for sonification and speech {ST1 – starting point 1, … , ST5 – starting point 5}

Responses to the survey were consistent with the quantitative results. 65% of subjects believed that they could find the target faster with sonification. Further, there was a slight tendency for subjects to consider it easier to navigate with sonification than speech. On a scale from Strongly Disagree (1) to Strongly Agree (5), 70% of subjects said it was easy to navigate with sonification, while 50% said it was easy to navigate with speech. However, the difference in mean scores (3.8 for sonification, 3.2 for speech) was not significant (t-test; n=20, df=38, t=1.58, p=0.12). When asked to directly choose which technique was easier to navigate with, subjects were evenly split. Finally, 65% of the subjects thought using sonification to navigate was more fun than using speech. This was consistent with remarks from several subjects that searching for a target was like playing a game: they even wanted to know how they performed in comparison to other subjects.

Subject responses to opened-ended questions on the survey shed more light on their preferences. A common theme was that sonification was quicker and required less cognitive effort:

I don't need to use my brain to think higher or lower number using non-speech system. I just feel the tone and switch to find the target.

Non-speech gave more rapid updates → allowed for a quicker response.

Non-speech is faster because there is less thinking involved… It's more natural because you are not comparing numbers.

On the other hand, some subjects praised the precision of speech:

Numeric distance was handy, beeps in non-speech audio pretty far apart.

Leaning towards speech, the distance indicator versus the tempo allowed me to take larger "steps" towards the target.

Easier to distinguish speech than to figure out where the beeps come from and what the beeps indicated.

6.2. Dual Task

In the dual task condition, subjects once again reached the target significantly faster with sonification. And the advantage increased: the mean time for sonification was 19.7 seconds, and for speech it was 32.9 seconds (t-test; n=100, df=198, t=4.32, p<0.01). Figure 4 shows the mean times for subjects to reach the target for the five different start configurations.

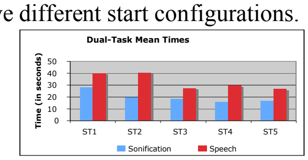

Figure 4. Mean time to target during the dual-task condition for sonification and speech

Compared to the single task condition, speech slowed down (from 26.9 seconds), while sonification actually got slightly faster (from 20.6 seconds). Neither change was significant, although the slowing down for speech neared significance (t-test; p=0.06), but the speed up with sonification did not (t-test; p=0.7). Figure 5 and Figure 6 detail the change for each of the 5 different start conditions. The increase in time for the speech condition isn't surprising. We expected that the additional cognitive processing required for the secondary task would reduce the amount of attention subjects could devote to the primary task of navigation, and thus would make it take longer. Further, it isn't surprising that speech was affected more than sonification, since the secondary task also consisted of spoken output that subjects had to attend to. It is surprising, however, that subjects actually sped up in the sonification condition. We discuss this further below.

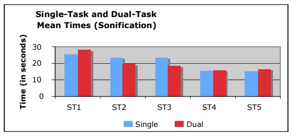

Figure 5. How the secondary task affected sonification performance

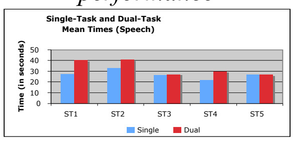

Figure 6. How the secondary task affected speech performance

Analysis of the subjects' letter responses again showed an advantage for sonification (Table 2). In the sonification condition, the subjects (collectively) heard a total of 318 "k k" sequences. They identified 266 of these sequences, or 83%. In the speech condition, the subjects (collectively) heard a total of 563 "k k" sequences. They identified 363 of these sequences, or 64%. Examining the performance of individual participants suggests that there are differences in the individual abilities for both sonification and sonification conditions. Nevertheless, it is not conclusive that individual differences have a stronger effect on one type of audio.

		Sonification			
	Total	**Mean**	**Min**	**Max**	**Std Dev**
Number of letter pairs heard	318	3	1	23	3.02
Correctly identified pairs	266	3	0	9	3.04
		Speech			
Number of letter pairs heard	563	6	1	22	4.4
Correctly identified pairs	363	4	0	14	2.75

Table 2: Letter responses statistics for sonification and speech conditions

Responses to the survey reinforced the quantitative results. First, subjects confirmed that the secondary task imposed increased cognitive demand. While 40% of subjects said that they had to concentrate hard during the sonification condition, 90% agreed with the corresponding statement for the speech condition. This difference is significant (t-test; n=20, df=38, t=4.07, p<0.01). Second, other responses showed that subjects perceived speech to impose a greater cognitive demand. Subjects agreed slightly with the statement that it was easy to listen for the letters in the sonification condition (3.3 out of 5). However, they disagreed with the corresponding statement for the speech condition (1.9 out of 5). This difference was significant (t-test; n=20, df=38, t=3.65, p<0.01). When asked directly whether it was easier to listen for letters while using sonification, 95% of subjects agreed. Finally, subjects found sonification less disruptive to their primary task. For sonification, subjects disagreed slightly with the statement that they had to pause to listen to the letters (2.5 out of 5). For speech, on the other hand, subjects agreed with the corresponding statement (3.8 out of 5). This difference was significant (t-test; n=20, df=38, t=3.46, p<0.01).

Subject comments reinforced the finding that sonification interfered less with the secondary task and imposed less cognitive demand.

With speech, a lot of interference hard to keep focused on finding the target.
I don't have to concentrate on understanding both things being said when doing the non-speech task.
I prefer non-speech because only had to listen to 1 "person" instead of two, I didn't have to pause, I "buffered".
I found myself distracted and off-task more often with speech.

6.3. Learning Effect?

We did additional analysis to try to understand why subjects performed slightly faster with sonification during the dual task condition than during the single task. Since they encountered the same sequence of five start conditions in each of the four main conditions, perhaps there was a learning effect; participants could have memorized the starting positions. We investigated this possibility by looking at the mean time subjects took to reach the target from each of the five starting points the first time they encountered it, the second time, the third time, and the fourth time. Figure shows the results.

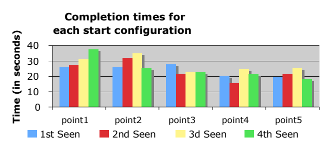

Figure 7. Time required to reach the target for each start position (point1, … , point2), over the four times it was encountered

The results do not show a clear effect, at least not one that is consistent across start positions. For the first and fourth start condition, subjects tended to slow down over time, while for the others there was at least some improvement in time. Note that this analysis does not separate speech from sonification. Thus, for each bar in the chart half of the data consist of times from subjects who were using speech and half from subjects using sonification.

Although, Figure 4 shows that there is a decreasing trend in the dual condition. Participants performed better in later trials, particularly in the sonification condition (Figure 5). All starting positions of the trials were at the same distance from the target but some were easier to find, for example, when the target was exactly on the right. Therefore, we will need to conduct a longer series of trials to better determine whether performance improves more quickly with sonification or with speech.

6.4. Summarizing the results: Answering the research questions

RQ1. Is sonification more effective than speech at delivering basic navigation information (distance and orientation to a target destination)?

Subjects found our sonification technique for indicating distance to the target – increasing and decreasing tempo and pitch – natural and easy to comprehend. Likewise, they found our simple spatialization of the tone effective in conveying orientation to the target. Most important, they navigated to the target more quickly using sonification than speech.

RQ2. Is sonification more effective than speech at delivering basic navigation information in the presence of a secondary (speech comprehension) task?

In the dual task condition, the performance advantage of sonification increased. Further, subjects made significantly fewer errors on the secondary task. We believe there are two reasons for this. First, processing speech imposes greater cognitive demand. Subjects note that it takes longer to comprehend speech messages than sonification, and speech messages require remembering and comparing numbers (distance to the target). Second, one speech stream interferes with the other, a point that subjects emphasized in their survey responses and reinforced with their comments. Some subjects did note that using a male voice for the navigation messages and a female voice for the sequence of letters did help them distinguish the two streams. However, in general, it was clear that sonification and speech interfered with each other much less. Therefore, we would recommend to designers of navigation aids that they consider carefully what types of environmental audio their users might be attending to, and use a different type of audio in their systems.

RQ3. For both single and dual task conditions, what are subjects' subjective preferences for speech and sonification?

In the single task condition, subjects believed (correctly) that sonification was faster, expressed some belief that sonification was easier, and thought that sonification was more fun. In their comments, some subjects noted that they preferred the precise absolute numbers speech gave them. Many others, however, said that the sonification output was simpler and faster to comprehend. For the dual task condition, subjects believed that speech made them concentrate harder, imposed more cognitive demand, and was more disruptive than sonification.

7. SUMMARY AND FUTURE WORK

We evaluated the use of two audio output techniques – speech and sonification – to deliver information necessary to navigate to a target destination. Effective information delivery is a fundamental issue that must be addressed to develop effective navigation aids. We evaluated these techniques in a 2D virtual space. We compared the two techniques, first in isolation, then in the presence of a secondary task (identifying repetitions in a sequence of spoken letters). In both cases, sonification was significantly faster, and its advantage increased in the presence of the secondary task. Subjects made fewer errors on the secondary task when using sonification. Finally, survey responses and comments showed that users generally preferred sonification.

Does this mean that we've shown that sonification is the right technique for delivering navigation assistance to both sighted and visually disabled people? Not yet. There are several clear limits to our study. Overcoming these limits suggests a number of avenues for interesting future work.

7.1. Testing with visually disabled users

None of our subjects were visually disabled. It's possible that visually disabled people might have different perceptions of and preferences for sonification and speech. Therefore, future studies should include both sighted and non-sighted participants.

7.2. Audio and visual information

We used an auditory (speech) secondary task to evaluate both sonification and speech. It would be interesting to repeat the evaluation of the two audio types when visual information is present. For example, (sighted) subjects can navigate to a target based on the audio while concurrently identifying repetitions in a sequence of images. Exploration of audio and speech together can reveal interesting issues related to using navigational systems while driving.

7.3. A more realistic environment

In our study, we used a very simple simulated space: there was only one target object and no obstacles. Our software already supports more complexity: obstacles can be added to the space, and we have designed sonifications to indicate their presence. Therefore, one possible follow-up study would be to replicate our study in a more complicated and realistic virtual space. This would put us in a much better position to move on to the real world, guiding subjects as they walk to a target destination.

7.4. Richer navigation information / integration of speech and sonification

We delivered only two pieces of information: distance to and orientation toward a target destination. With so little information, the properties of sonification – quickness, ease of interpretation, ease of conveying dynamically changing data – were distinct advantages. However, as more – and more complicated – types of information must be communicated, the flexibility and generality of speech will become increasingly advantageous. For example, perhaps people will want to be notified about different types of objects in the environment in addition to their destination: information booths, rest rooms, drinking fountains, etc. Of course, non-speech audio can be used for this purpose; "audio icons" or natural sounds – e.g., recordings of water running – can be used. However, eventually this approach begins to reach a limit. It gets hard to identify appropriate natural sounds or audio icons that people can readily distinguish, so users have to start learning and memorizing the meaning of new types of sounds. At this point, it makes sense to turn to speech, since users already understand words that can be used to convey the information in question.

7.5. Other output modalities

Speech and sonification both are auditory. Thus, to some extent, either one will interfere with a person's ability to attend to ambient noises in the environment or to carry out a conversation. As we've mentioned earlier, there has been some work done on tactile or haptic output. We consider this a promising avenue of research. We think tactile output will be well suited for the same types of information as sonification. However, this is just a conjecture, and it would be well worth testing. Moreover, the multiple resource theory of attention **Error! Reference source not found.** suggests that tactile output may have an advantage over sonification. This theory suggests that people have different perceptual/cognitive capabilities, and that attending to data received via one sense (e.g., hearing) will not interfere with data received via another (e.g., touch). This predicates that tactile information delivery would be even less distracting than sonification when other audio sources must be attended to.

In conclusion, we have addressed a problem of great practical interest: developing navigational aids to bring greater independence to visually disabled people and decrease cognitive load for both sighted and non-sighted users when several tasks are involved. We've explored one sub-problem: how navigation information can be delivered effectively, and with minimal distraction. Our results show that, in some circumstances, sonification is quite effective: subjects perform better than with speech, and prefer sonification, too. Finally, we identified a set of research challenges that must be addressed to realize the promise of technological navigation aids.

REFERENCES

[1] A speech synthesizer written entirely in the JavaTM programming language. http://freetts.sourceforge.net/docs/index.php.

[2] Baig, C., "Where in the world am I? Your phone might know." USA TODAY (June 2006), http://www.usatoday.com/tech/products/services/2006-06-01-location_x.htm.

[3] Barrass, S., Kramer, G. Using Sonification. Multimedia Systems, Springer-Verlag New York, Inc (1999), 23-31.

[4] Bradley, N.A., Dunlop, M.D. An Experimental Investigation into Wayfinding Directions for Visually Impaired People. Personal and Ubiquitous Computing '05, ACM Press (2005), 395-403.

[5] Fitch, W.T., Kramer, G. Sonifying the Body Electric: Superiority of an Auditory over a Visual Display in a Complex, Multivariate System, In Auditory Display: Sonification, Audification and Auditory Interfaces, edited by Kramer. Proc. SFI Studies in the Sciences of Complexity, Addison-Wesley (1994), 307-326.

[6] Gaver, W. Synthesizing Auditory Icons. Proc. INTERCHI '93, ACM Press (1993), 228-235.

[7] Gollege, R.G., Martson, J.R., Loomis, J.M., Klatzky, R.L. Stated Preferences for Components of a Personal Guidance System for Nonvisual Navigation. Journal of Visual Impairment & Blindness, JVIB (2004), 135-147.

[8] Java Foundation Classes (JFC/Swing). http://java.sun.com/products/jfc/.

[9] Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., Neuhoff, J. Sonification Report: Status of the Field and Research Agenda. Prepared for the National Science Foundation by Members of the International Community for Auditory Display Editorial Committee and Co-Authors. (1997), http://www.icad.org/websiteV2.0/References/nsf.html.

[10] Kulyukin, V., Gharpure, C., De Graw, N., Nicholson, J., Pavithran, S. A Robotic Guide for the Visually Impaired in Indoor Environments. Proc. RESNA '04, http://www.cc.usu.edu/~cpg/pubs/Resna04.pdf.

[11] Lai, J., Cheng, K., Green, P., Tsimhoni, O. On the Road and on the Web? Comprehension of synthetic and human speech while driving. Proc. SIGCHI '01, ACM Press (2001), 206-212.

[12] Loomis, J.M., Marston, J.R., Gollege, R.G., Klatzky, R.L. Personal Guidance System for People with Visual Impairment: A Comparison of Spatial Displays for Route Guidance. Journal of Visual Impairment & Blindness, JVIB (2005), 219-232.

[13] Loomis, J.M., Golledge, R.G., Klatzky, R.L. Navigation System for the Blind: Auditory Display Modes and Guidance. Journal of Presence, (1998), 193-203.

[14] Ludford, P.J., Frankowski, D., Reily, K., Wilms, K., Terveen, L. Because I carry my cell phone anyway: functional location-based reminder applications. Proc. SIGCHI '06, ACM Press (2006), 889-898.

[15] MacLean, K. Designing with Haptic Feedback. Proc. ICRA'00, IEEE Press (2000), 783-788.

[16] Matta, S., Rudolph, H., Kumar, D.K. Haptic Eyes: Mapping Visual Information in to Haptic Displays. Proc. SPIE '05, SPIE (2005), 1354-1362.

[17] Okamoto, M., Junichi, A., Ito, K., Ono, T., Komatsu, O. CyARM: an Alternative Aid Device for Blind Persons. Ext. Abstracts CHI '05, ACM Press (2005), 1483-1488.

[18] Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. Human Brain Mapping, PubMed (2005), 46-59.

[19] Package javax.sound.sampled. http://java.sun.com/j2se/1.4.2/docs/api/javax/sound/sampled/package-summary.html.

[20] Pitt, J., Edwards, A.D.N. Improving the usability of speech-based interfaces for blind users. Proc. SIGACCESS '96, ACM Press (1996), 124-130.

[21] Steven, K., Brennan, D., Parker, S. Simultaneous Manipulation of Parameters of Auditory Icons to convey Direction, Size, and Distance: Effects on Recognition and Interpretation. Proc. ICAD '04, (2004).

[22] Stevens, C., Brennan, D., Parker, S. Marketbuzz: Sonification of Real-Time Financial Data. Proc. ICAD '04, (2004).

[23] Strothotte, T., Fritz, S., Michel, R., Raab, A., Petrie, H., Johnson, V., Reichert, L., Schalt, A. Development of Dialogue Systems for the Mobility Aid for Blind People: Initial Design and Usability Testing. Proc. ASSETS '96, ACM Press (1996), 139-144.

[24] Vilimec, R., Hempel, T. Effects of Speech and Non-Speech Sounds on Short-Term Memory and Possible Implications for In-Vehicle Use. Proc. ICAD '05, (2005).

[25] Walker, B.N., Lane, D.M. Psychological Scaling of Sonification Mapping: a Comparison of Visually Impaired and Sighted Listeners. Proc. of ICAD '01, (2001).

[26] Walker, B.N., Kramer, G. Mappings and Methaphors in Auditory Displays: An experimental assessment. Proc.