# Reading On–the–Go: A Comparison of Audio and Hand–held Displays

Kristin Vadas, Nirmal Patel, Kent Lyons, Thad Starner and Julie Jacko
College of Computing and Department of Biomedical Engineering
Georgia Institute of Technology
Atlanta, GA 30332
{vadas,merik,kent,thad}@cc.gatech.edu
jacko@bme.gatech.edu

## ABSTRACT

In this paper we present a 20–participant controlled experiment to evaluate and compare a head–down visual display and a synthesized speech audio display for comprehending text while mobile. Participants completed reading comprehension trials while walking a path and sitting. We examine overall performance and perceived workload for four conditions: audio–walking, audio–sitting, visual–walking, and visual–sitting. Results suggest audio is an acceptable modality for mobile comprehension of text. Participants' comprehension scores for the audio–walking condition were comparable to the scores for the visual–walking condition. More importantly, participants saw improvements in their ability to navigate the environment when using the audio display.

**Categories and Subject Descriptors:** H.5.2 [User Interfaces]: [Evaluation/methodology; Screen design (e.g., text, graphics, color); Auditory (non-speech) feedback]

**General Terms:** Human Factors

## 1. INTRODUCTION

There has been an explosion in the use of mobile devices such as mobile phones, PDAs, smartphones, laptops, palmtops and wearables. In 2004 there were 1.3 billion mobile phone subscribers, and two billion are predicted by 2007 [2]. Wireless text messaging has become widespread, with predictions that soon over one trillion messages will be sent per year [14]. Whether browsing the web on a smartphone while waiting in line, trying to find a friend's telephone number while walking to a restaurant, or reading through a text–message from a colleague while hurrying to a meeting, the ability to read on–the–go is quickly becoming an important skill.

As reading on–the–go becomes increasingly common, we realize our ability to read while walking is limited. Reading on–the–go involves managing two main tasks in parallel: comprehending the text in question and navigating the environment. When using a traditional visual display, the user must split visual resources between viewing the environment and reading text on a screen. The limits in our ability to efficiently navigate the environment while reading for comprehension may be largely attributed to inherent physical and cognitive constraints. However, our ability to reach these limits is likely confounded by the constraints of the devices we use. Text size, instability of displays, and the heads–down nature of most of today's mobile displays are just some examples of the design features that make reading on such devices while in motion difficult.

After completing initial work to evaluate several different visual display types for reading while waking [18], we became interested in exploring if a different modality might better support mobile comprehension. Audio offers a hands–free, eyes-free alternative to visual displays. In this paper we present a controlled laboratory experiment comparing a synthesized speech interface and a visual head–down interface (Figure 1) for comprehension of text while walking. We explore the tradeoffs between using visual and audio interfaces for mobile comprehension of text by discussing our experimental findings.

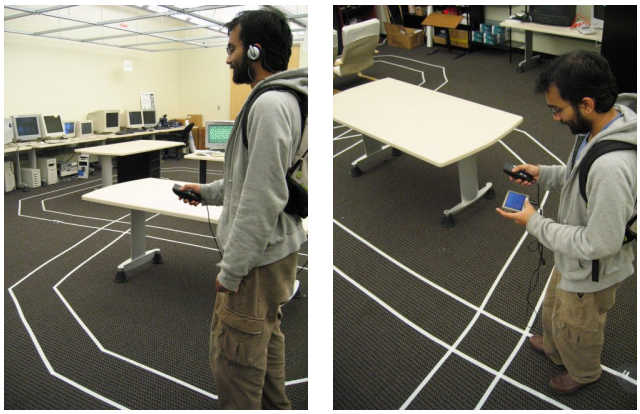

**Figure 1: Audio–walking condition (left): Participants walk around the path while listening to synthesized speech through headphones. Visual–walking condition (right): Participants walk around the path while reading text on a small visual display.**

## 2. RELATED WORK

Reading is a fundamental task performed on mobile devices. Mustonen *et al*. evaluated legibility of text on mobile phones while walking at different speeds, both on a treadmill and down an empty

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
*MobileHCI'06* September 12–15, 2006, Helsinki, Finland.
Copyright 2006 ACM 1-59593-390-5/06/0009 ...$5.00.

corridor [15]. They found visual performance deteriorates with increased walking speed and that, as subjective task load increases, performance declines. Experiments conducted by Barnard *et al*. [3, 4] saw similar results. In these studies, participants completed word search and reading comprehension tasks on a PDA while either walking on a treadmill, following a path on the floor, or sitting. Participants rated subjective workload higher while walking on a path as opposed to walking on a treadmill. Additionally, participants read faster, had better comprehension scores, and perceived less workload while sitting as compared to walking on a path.

In addition to work exploring mobile reading, there has been some work exploring the use of speech as a communication medium for mobile systems. Several such systems use synthetic speech, often in combination with another output source. Nomadic Radio is an all audio output wearable computer that provides numerous services such as a calendar, email and news access [17]. All information is relayed to the user through a synthetic voice or pre-recorded sound files. A more recent system, MATCH, uses a combination of synthetic speech and a graphical display for output [8]. Other systems have used pre–recorded natural speech instead of synthesized speech. NewsComm is a mobile, hand-held system which allows a user to index pre–recorded news broadcasts [16]. In this work, Roy *et al*. explored different ways of structuring and navigating the audio, one of which involved annotating the audio at semantically significant points using pause and pitch. While these systems utilize speech output, little work has been done evaluating the effectiveness of synthesized speech for text comprehension while mobile.

There have been several past studies evaluating the effectiveness of synthetic speech in comparison with natural human speech. Lai *et al*. measured the effects of various task conditions in the comprehension of synthetic speech [10]. They used a variety of passages ranging from short reminders to spoken email and news. In one condition the mean accuracy declined as the passage became longer. A second study by Lai *et al*. examined the comprehensibility of synthetic speech while driving [9]. The experiment was conducted in a driving simulator, and messages of various lengths, such as navigational cues, email snippets or news stories were used. While participants rated the synthetic speech lower, they found that voice type had no effect on driving performance. Interestingly, they found an increase in driving workload led to an increase in performance, even on questions regarding longer news stories. This result may be due to an increase in overall focus and attention when the task became difficult.

## 3. VISUAL DISPLAYS AND READING ON-THE-GO

In previous work, we examined the use of different visual displays for reading while walking [18]. Our findings motivated us to explore audio as an alternative for mobile text. We briefly discuss the experiment and related findings.

Interested in the problem of reading while on–the-go, we explored how different types of display technologies might influence reading comprehension while walking. We were particularly interested in the possibility of using head-mounted displays, such as those used by wearable computer users [13]. Thus, for the study, we chose three display devices, each with different design features: a MicroOptical head–mounted display (Figure 2), a Sony electronic ink e-book reader (Figure 3), and an OQO palmtop computer (Figure 4). We chose the head–mounted display because it allows head–up and hands–free use. The OQO served as

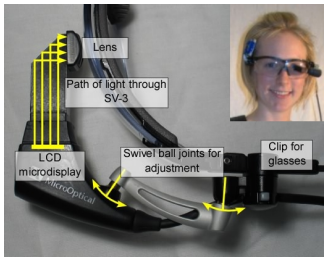


**Figure 2: Top-down view of the MicroOptical SV-3 head-mounted display.**

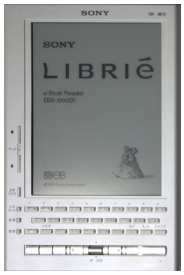


**Figure 3: The Sony e-book reader.**

**Figure 4: The OQO Model 01 palmtop computer.**

a representative of hand-held devices typical in today's market, having high resolution but suffering from issues such as glare [4]. Finally, we chose the Sony e-ink device for its novel, low power reflective electronic ink technology that provides for a larger range of viewing angles and minimal glare.

Our in–lab experiment was a single-variable within-subjects design with one condition per device. For each condition, participants completed sets of ten reading comprehension trials while walking around a path. The path was approximately 40 meters long and 30 cm wide. Each comprehension trial consisted of a reading a passage followed by two related multiple choice questions. Participants were instructed to stay inside the lines of the path and to continue walking until the set of ten trials was complete. Additionally, they were asked to answer the questions as accurately as possible. To assess participant performance, we recorded reading time, response accuracy, path accuracy, and walking speed. After each condition, participants completed the NASA Task Load Index (NASA-TLX) survey [6, 7] to provide an assessment of their perceived workload for each device.

Our hypothesis was that the head–mounted display would allow the participants to more easily monitor their environment while reading relative to the head–down alternatives. However, our data and analysis of the NASA-TLX results showed participants found the head–mounted display the *most* difficult to use. When there were statistically significant differences in the scores, both of the hand–held displays yielded lower perceived workload than the head–mounted display. A repeated measures ANOVA on the display condition shows no statistical difference in ability to answer the questions correctly; the average accuracy across all conditions was 69.4%. We did find significant pairwise effects for reading time. Results showed participants read faster on both of the hand–held devices than on the head–mounted display. This finding is reinforced by participant comments stating they would often lose their place in the text while reading on the head-mounted display. Several participants indicated that they lost their place due to motion of the head–mounted display, while others mentioned being distracted by the environment. The issue of distracting environmental backgrounds is consistent with findings from studies of stationary head-mounted display use [11].

Overall, walking performance results were poor, regardless of display type. For the normalized measures of average speed and number of steps off the path, there were no statistically significant differences between the three different displays. When reading while walking, the participants slowed their walking rate from an average natural walking speed of 1.01 m/s to 0.69 m/s during the trials. Many participants commented that they were surprised by the difficulty of completing the reading comprehension tasks while walking, regardless of display type. We anecdotally noted that participants' gates were inconsistent throughout the trials. Additionally, we noted that participants had difficulty staying inside the path, frequently stepping on or over the lines.

These results led us to question if using a visual display is the best option for presenting text for comprehension while walking. In this paper, we present a follow–up study in which we used a similar method to evaluate an audio interface for "reading" while walking. Since audio does not require visual resources, we hypothesized introducing audio output would result in less resource contention [5] and allow participants to more easily and effectively process and navigate their environment (i.e. follow the path more accurately, walk faster, etc.). However, we realize audio is not without its limitations. Audio is inherently linear and thus enforces sequential parsing of the text presented. We expected our participants would spend more time listing to the audio output of the system relative to time spent reading text on a screen. Additionally, since we chose to use computer synthesized speech, we expected that comprehension accuracy would be less for our audio conditions relative to reading the text on the visual display.

# 4. METHOD

To explore the effectiveness of synthesized speech as a mobile audio display we examine two independent variables. The first variable is display type (audio and visual) and the second is mobility (sitting and walking). The sitting condition allows for a base–line assessment of each participants' comprehension level, while the walking condition allows for comparison of the audio and visual displays for use in motion. Our study is a within subject 2x2 Latin square experimental design with four conditions: audio–walking, audio–sitting, visual–walking, and visual–sitting. This design is largely based on the studies performed by Barnard *et al*. [3, 4] where participants walk a predefined path while performing reading comprehension trials.

## 4.1 Experimental Trials and Conditions

### *4.1.1 Comprehension Trials*

To assess the ability of our participants to comprehend text in the various conditions, we selected a task which involved reading or listening to a short passage and then answering two multiple choice questions based on the passage. Both the passages and questions were selected to be short enough to fit on one screen without scrolling, and were taken from a book designed to prepare high school students for standardized tests [12] (the same source used in the Barnard et al. experiments). The passages, which range from one to three paragraphs and average 107 words long, are composed of both fictional stories and non–fictional messages. The audio version of the passages are on average 42.2 seconds long. We consider each combination of a passage and the two related questions to be a trial. Participants completed five trials for each condition, resulting in a total of 20 passages and 40 questions across all four conditions. The same 20 trials were used for all participants, but the order and distribution of the trials across conditions was randomized for each participant to minimize ordering effects. Additionally, not all of the trials are of exactly the same difficulty, thus randomizing the distribution of the trials across the conditions limits trial difficulty as a confounding factor.

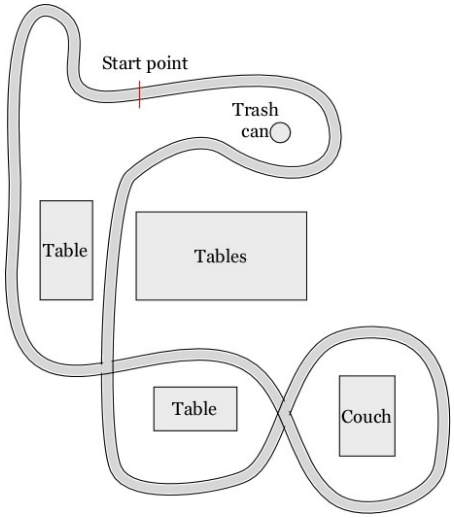


**Figure 5: The path participants walked along.**

### *4.1.2 Mobility: Sitting and Walking*

For the sitting condition, participants sat at a table in the laboratory and were instructed to sit however they felt comfortable. Participants were reminded they should not get up until they finished all five trials. For the walking part of the experiment, participants followed a path, approximately 46 meters long and 30 cm wide taped on the floor in a laboratory environment (Figure 5). They were told they could slow down or speed up, so long as they did not stop until they completed all five trials for the condition. The experimenter reminded the participants to stay inside the lines of the path as best as possible. The path curved and required the participants to navigate around several objects, such as tables of varying heights (Figure 5). Both the path and positioning of obstacles remained constant across all participants. The path was marked at the starting point and at 30.5cm (1 foot) intervals with pencil (barely visible to participants) to facilitate measuring distance. As with the Barnard et al. studies [3, 4], the direction the participants walked on the path (clockwise or counter-clockwise) was randomized across conditions and participants to help minimize learning effects.

### *4.1.3 Display: Audio and Visual*

For the audio display, participants wore a pair of head–phones and the trials were presented using synthesized speech. The OQO screen was always kept out of the participant's view, either on the opposite side of the desk for the sitting condition or in a backpack for the walking condition. For the visual display conditions, participants held the OQO in their non–dominant hand. The input device for responding to the questions was held in their dominant hand for all conditions.

## 4.2 Equipment and Software

### *4.2.1 Base Platform: OQO*

The OQO Model 01 was the base platform for the experiment, hosting the software, logging the data, and serving as the display for the visual conditions and the source of audio for the audio

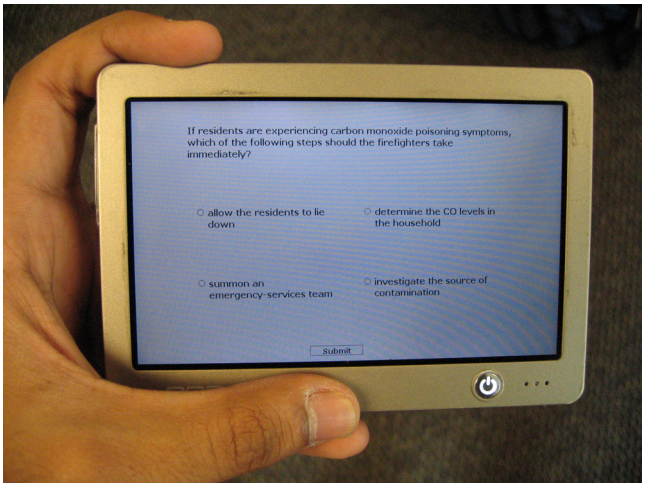


**Figure 6: The OQO Model 01 palmtop computer is used as the base platform to run the software and collect user input, and is also used as the display in the visual display condition.**

conditions. The OQO (Figure 6) is a small form–factor palmtop computer. It weighs approximately 400g and fits comfortably in the palm of the hand. The OQO display is a transflective TFT liquid crystal display (LCD) that measures 109.5x66.6mm and has a resolution of 800x480, resulting in approximately 185dpi. The OQO has a Transmeta Crusoe 1Ghz processor, 256MB memory, and a 20GB hard drive. It also has a variety of peripheral ports including USB 1.1 host, VGA output, and Firewire. The screen on the OQO slides to reveal a miniature QWERTY keyboard; however, during our study the keyboard remained hidden.

### *4.2.2 Software*

Our custom experimental software, written in Java, presented the comprehension trials and logged all of the data and participant input. All user input was time-stamped by the software (in milliseconds) and logged at the button press level.

For any given trial, the software progresses through three states (Figure 7): passage (p), question one (q1), and question two (q2). State transitions are initiated by button presses on the customized keypad (Figure 10). In the passage state, the software shows a text passage (Figure 8) or plays a synthesized speech reading of the text depending on the display condition. In the audio condition, the rewind or fast–forward button may be pressed at any time to move backward or forward through the passage speech. Pushing the submit button while in the passage state (p) moves the software into the question one state (q1). Once in q1, there is no way to navigate back to the passage. The software then displays the first question along with the answers (Figure 9) for the visual condition. In the audio condition, the question speech is automatically played, but the answers are not. As with the passage, the question speech may be reviewed with the rewind and fast–forward buttons. Pressing one of the answer buttons on the input device will select the answer and play the corresponding speech. The participant must press each of the answer buttons to hear all of the potential answers. For both conditions, pressing the submit button moves the software into the question two state (q2) and there is no way to return back to question one. Interaction works the same as in the question one state. Finally, pressing the submit button moves the software back to the passage state (and on to the next trial).

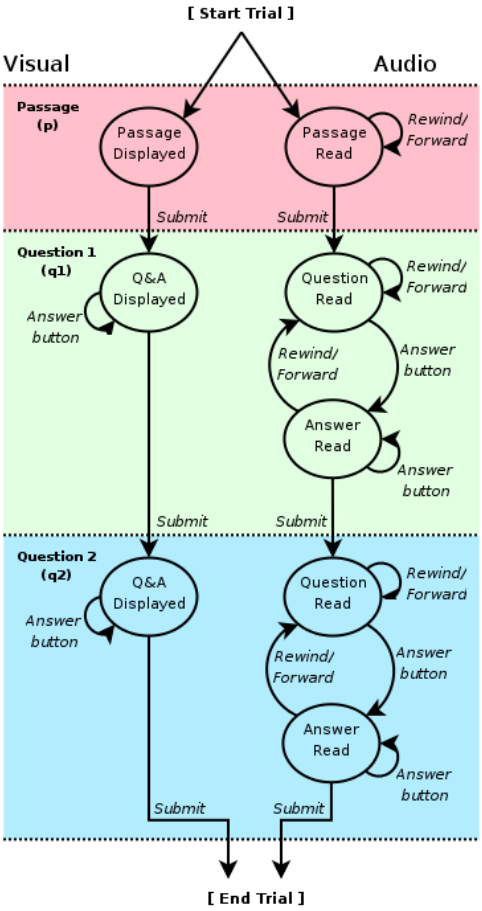


**Figure 7: This software state diagram describes the interaction flow through the comprehension trials for both the audio and visual interfaces.**

The audio display plays pre–recorded synthetic voice readings of passages, questions, and answers. While human speech has proven a more effective output method [10], we explicitly chose to explore synthetic speech. Text–to–speech generators of reasonable quality are in existence today, and thus text–to–speech presents a technologically feasible approach to incorporating real–time speech output of text in mobile devices. The synthetic voice used in our study is the "Crystal" (female, US English) model of AT&T's Natural Voices package. The AT&T system was used because of the level of control available in selecting rate of speech, pronunciation and pausing. Our software allows participants to rewind and fast–forward through the passages at the sentence level. Sentence level pauses were chosen because they "suggest the boundaries of material to be analyzed and provide vital cognitive processing time" [1]. We used non–speech sound cues to provide feedback for actions or state transitions. Rewinding is associated with a falling pitch and forwarding is associated with a rising pitch. If a user attempts to move beyond the beginning or end of an audio segment a cartoon "boing" is played. These effects follow the interface used in SpeechSkimmer [1]. In addition, a low, short tone indicates that the application has finished speaking. Finally, a short clicking sound plays whenever the user pushes the submit or done button. The volume of the non-speech cues is kept lower than the volume of the speech to make them less obtrusive.

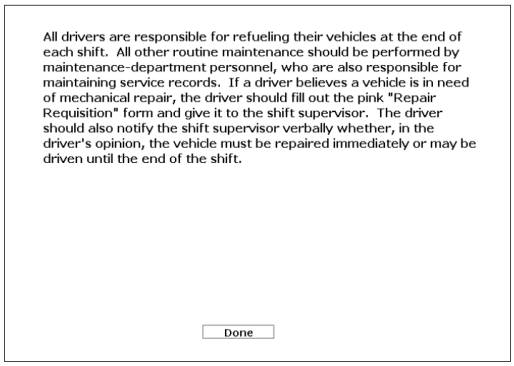


**Figure 8: An example reading passage.**

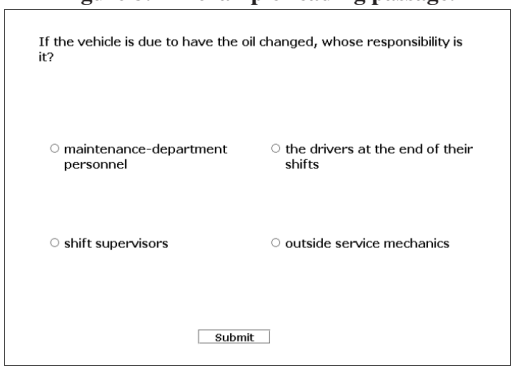


**Figure 9: A multiple choice question example.**

### *4.2.3 Input Device*

The custom input device used in this experiment is a modified one–handed Twiddler keyboard with seven buttons (Figure 10). The four buttons at the bottom spatially correspond to answers in the reading comprehension task, and the one central button is red and conceptually corresponds to a "Done" or "Submit" key (Figures 8 and 9). Two buttons at the top are used in the audio condition for skipping backwards (left) or forward (right) at the sentence level. In the visual condition, these buttons have no effect. In contrast to the Barnard *et al*. studies [3, 4], we chose to use a separate keypad for input because we are interested in only studying the effects of mobile output. By standardizing the input device across the conditions, we removed input as a potential source of differing influence.

## 4.3 Dependent Measures

To assess participant performance, we recorded reading time, response accuracy, path accuracy, and walking speed. Reading time is the time from when a passage is first displayed to when the participant presses the red submit button to proceed to the question. Response accuracy is whether or not the participant selected the correct answer. All of these values were calculated from the data logs after the experiment was completed. Total distance was calculated by counting the number of laps (full and partial) around the path. Path accuracy is the number of times the participant stepped on or outside the lines of the path, normalized by the total distance (in meters) traveled on the path. Finally, average walking speed is the total distance walked divided by the total time to complete all five trials.

To assess perceived workload, each participant completed the standard NASA Task Load Index (NASA-TLX) scale and demand

**Figure 10: The modified Twiddler keypad used for input.**

comparison surveys after each display condition (upon completion of the corresponding set of five trials). The NASA-TLX is a questionnaire used to measure subjective workload ratings. Previous studies have indicated that it is both a reliable and valid measure of the workload imposed by a task [6, 7]. The NASA-TLX consists of six scales: mental demand, physical demand, temporal demand, performance, effort, and frustration; each scale has 21 gradations. For each scale, individuals rate the demand imposed by the task. In addition, they rank each scale's contribution to the total workload by completing 15 pairwise comparisons between each combination of scales. The overall workload rating is calculated by summing the product of each scale's rating and weight. This calculation results in a score between 0 and 100. It reflects an individual's perception of the amount of workload devoted to each of the scales, along with each scale's contribution to overall workload [7].

## 4.4 Procedure

The experiment began for each participant with a brief description of the experiment, an introduction to the NASA-TLX questionnaire and a short background survey.

### *4.4.1 Training*

Next there was a training session composed of four trials, two with each display type. The training session was designed to instruct the participant how to use the interfaces and to clarify any questions about the comprehension trials. First, the experimenter showed the participant how to use the visual display. The experimenter stepped through the first training trial, explaining when to press each button. The participant went through the second training trial on their own but were allowed to ask the experimenter any questions if desired. Next, the experimenter went through the same procedure with the audio interface using a set of small speakers and again the participant performed the second example on their own. Finally, at the beginning of each of the four conditions (audio–walking, audio–sitting, visual–walking, and visual–sitting), the participant completed one additional training trial in the respective mode (i.e. for the audio walking condition, participants completed a practice trial with the audio interface while walking around the track).

### *4.4.2 Natural Walking*

Baseline data for the "natural" walking speed along the path and number of steps off the path were collected at the beginning of the experiment (immediately after training) and at the end of

the experiment (after the final condition). The participant was instructed to walk once around the path in each direction at a comfortable pace while trying to stay inside of the path boundaries. The time to complete each lap was recorded, as well as the number of steps on or over the lines demarcating path.

#### *4.4.3 Trials*

At the beginning of each condition, the researcher configured the software and hardware as needed. The experimenter reminded the participant they would complete five trials in a row and asked the participant to answer the questions about each passage as accurately as possible.

For the walking conditions, the participants were instructed to continue walking until they finished the fifth trial and were asked not to stop in–between trials. The experimenter informed them that they could slow down or speed up as desired, but should not stop until all five trials were completed. As the participant walked around the path and completed the trials, the researcher followed behind as quietly as possible and used a tally counter to track of the number completed laps, as well as the number of times the participant stepped on or outside the lines of the path. When the participant completed the final trial, the software informed the participant to stop and the experimenter recorded the participant's final position.

At the end of each condition, the experimenter directed the participant to complete the NASA-TLX survey, reminding them to consider only the most recent five trials. After the TLX, the procedure was repeated with the remaining display and motion conditions. Finally, at the end of the study, the researcher asked the participants to share any comments they had about their experience with any of the displays and the task performed.

### 4.5 Participants

Twenty-six participants were recruited from the student body by word-of-mouth. We did not control for any demographic factors (i.e. gender, eye-sight, native language, etc.). All participants were either compensated $10 per hour or received one extra-credit point for a class they were taking regardless of their performance. Time to complete the study ranged from 38 minutes to 68 minutes. Of the 26 data sets generated, only 20 data sets contained all of the information needed for the study. Technical difficulties with the experimental hardware (mainly a result of the system overheating which caused the software to freeze during the experiment) resulted in six incomplete data sets. We consider only the 20 complete data sets throughout the rest of this paper.

The 20 participants ranged in age from 18 to 29 years, with a median of 22 years. One participant was left-handed, eighteen were right-handed, and one was ambidextrous. The three non-native English reading/speaking participants had experience reading and speaking English that ranged from 8 to 15 years. 17 of the participants were male and 3 were female.

## 5. RESULTS

Our 20 participants read a combined total of 400 passages and answered 800 questions. Table 1 shows the percentage of questions answered correctly for each condition. An analysis of variance (ANOVA) only shows a main effect for mobility ($F = 12.5, p < 0.001$). Not surprisingly, the data show that the participants answered the questions more accurately (had higher comprehension scores) while stationary (M=81.5%, SD=16.1%) than while mobile (M=67.5%, SD=18.9%).

Table 2 shows the average time spent reading or listening to each passage. An ANOVA reveals a main effect for display type

| | Audio | Visual | Mobility Means |
|---|---|---|---|
| Walking | 65.0% (21.4) | 70.0% (16.2) | 67.5% (18.9) |
| Sitting | 81.0% (17.7) | 82.0% (14.7) | 81.5% (16.1) |
| Display Means | 73.0% (21.0) | 76.0% (16.5) | 74.5% (18.8) |

**Table 1: Mean percent correct for each condition with standard deviations.**

($F = 23.4, p < 0.001$). As hypothesized, the participants took longer listing to the passages (M=53.1s, SD=12.0s) than reading them (M=39.4s, SD=13.6s).

| | Audio | Visual | Mobility Means |
|---|---|---|---|
| Walking | 52.82s (8.66) | 43.13s (14.6) | 47.98s (12.82) |
| Sitting | 53.46s (14.87) | 35.63s (11.67) | 44.54s (15.98) |
| Display Means | 53.14s (12.01) | 39.38s (13.59) | 46.26s (14.5) |

**Table 2: Mean time to read or listen to each passage for each condition with standard deviations.**

The average length of a spoken passage is 42.17 seconds with a standard deviation of 15.87. There is approximately an 11 second disparity between the average time spent listening to the passage (M=53.14s, SD=12.01) and the actual length of the passage. Approximately 4 seconds (M=4.34, SD=2.12) of this extra time was spent doing nothing while the remaining time was spent re-listening or navigating through the passage. There was no statistical difference in the extra time between listening while walking or sitting.

Next, we analyze the overall workload ratings (Table 3). An ANOVA reveals a main effect for mobility ($F = 29.3, p < 0.001$) as well as an interaction effect ($F = 7.7, p < 0.01$). The main effect indicates that participants had a higher workload while walking the path (M=54.1, SD=16.9) than while sitting (M=35.41, SD=15.2). Examining the table, the interaction can be seen in the visual condition with the stationary condition rated as having the least workload (M=30.9, SD=14.7), while the walking condition was rated highest (M=59.2, SD=16.5). The audio condition also shows a similar but smaller increase from sitting (M=39.9 SD=14.6) to walking (M=49.1, SD=16.0).

| | Audio | Visual | Mobility Means |
|---|---|---|---|
| Walking | 49.05 (16.03) | 59.23 (16.54) | 54.14 (16.88) |
| Sitting | 39.9 (14.6) | 30.92 (14.69) | 35.41 (15.15) |
| Display Means | 44.48 (15.83) | 45.08 (21.07) | 44.78 (18.52) |

**Table 3: Mean total TLX workload for each condition with standard deviations.**

### 5.1 Walking Performance

As we were most interested in the comparison of our participants' ability to comprehend the text while mobile, we next

examine the differences between the visual–walking and audio–walking conditions in more detail. A Student's t-test reveals no statistical difference (p=0.35) for comprehension accuracy (the percentage of questions answered correctly) between the audio–walking and visual–walking conditions (Table 1). In contrast, there is a statistically significant difference for the time spent reading/listening to the passages ($p<0.05$, Table 2). Supporting the overall ANOVA results, we found participants spent longer listening to the synthesized speech than reading the text on the handheld display while walking. We also found a statistically significant difference in the perceived total workload as measured by the NASA TLX (Table 3), with audio being rated more favorably than reading ($p<0.05$).

We also tracked measures related to the walking portion of the task. In particular, we recorded the walking speed and number of steps off the path per meter during the visual and audio conditions. The mean values for each condition are shown in Table 4.

| | Audio | Visual | Natural |
|---|---|---|---|
| Speed (m/s) | 1.03 (0.21) | 0.91 (0.14) | 1.20 (0.17) |
| Off-steps / m | 0.02 (0.03) | 0.09 (0.07) | 0.03 (0.03) |

**Table 4: Mean speeds and off–steps for the audio, visual, and base–line walking with standard deviations.**

A one way ANOVA reveals statistically significant differences between the listening, reading and natural walking speeds ($F = 14.1, p < 0.001$). A post–hoc analysis using a paired Student's t-test indicates statistically significant differences at the p=0.001 level between each pair of variables with the average natural (base–line) speed being the fastest (M=1.20m/s, SD=0.17), followed by the average audio–walking speed (M=1.03m/s, SD=0.21), and finally the visual–walking speed (M=0.91m/s, SD=0.14).

There is also a statistically significant difference between the number of steps off the path between the audio, visual and the natural walking conditions ($F = 13.0, p < 0.001$). Post–hoc pairwise analysis indicates statistically significant differences between the audio and visual conditions and between the visual and natural conditions; however, there is no statistically significant difference between audio and natural.

# 6. DISCUSSION

Overall, these data show that the participants performed well using the synthesized speech audio display while in motion. Particularly, the results support our hypothesis that introducing audio output would allow participants to more easily and effectively process and navigate their environment by freeing up visual resources. Our findings suggest that having an audio output option would be a beneficial and useful feature for mobile devices that involve the presentation of text passages.

Examining participant walking performance, we found statistically significant differences in favor of the audio condition. Walking speed and path accuracy were higher in the audio condition than in the visual condition. While reading, the participants took many more steps off the path than with either the audio display or when walking naturally without a task. Anecdotally, we also noted that most participants' gaits were inconsistent while using the visual display. Participants often stumbled and fluctuated between really small quick steps and larger slower steps. In contrast, gait was typically more consistent in the audio condition.

Another important factor in favor of using an audio display for mobile comprehension is the rating of workload. With our past experiment comparing three visual displays, we anticipated participants having difficulty with our comprehension task. Furthermore, we speculated that the synthesized audio would have drawbacks, such as being hard to understand or difficult to use because of audio's linear nature, that might negatively impact the participants' experience. Unexpectedly, our participants subjectively rated listening to the audio display as less demanding than reading the visual display while walking.

Our metric for comprehension performance is the ability of participants to correctly answer the comprehension questions. The only statistically significant result that we found is that participants were more accurate while stationary than while mobile, which was expected. The task of navigating the environment requires attentional and working memory resources that can no longer be devoted to the comprehension of the passages when the user is walking the path [5].

The data on reading performance reveals that, in the walking conditions, our participants spent more time on average listening to the passage (M=52.82s, SD=8.66) compared to when they read the text themselves (M=43.13s, SD=14.6). The longer time for audio is not surprising since audio is inherently linear and the participants needed to listen to the speech at the pace established by the system. Furthermore, the linear nature of speech makes it difficult to scan the passage, whereas reading text affords looking a few lines farther forward or backward to quickly review information. To overcome the linear nature, we added the ability to skip forward and backwards in the text; the data shows that the participants spent approximately 7s in a passage using these features.

While the audio display was slower than the visual display, it is important to understand the impact of this result. Since we are looking specifically at use of mobile devices while on-the-go, it is possible that speed may not be the most important factor. Depending on the application, the user may be filling time that would otherwise be wasted (i.e. they may be multi-tasking between walking and texting as opposed to just walking). Coupled with the TLX results on workload, audio might offer a good balance of performance and functionality for a given amount of effort. Together these results imply a class of applications that, instead of optimizing for time efficiency, allow for better use of our spare capacity when mobile.

The overall performance for audio while mobile is likely a result of several factors. First, the audio display we used required only one hand (to use the input device), whereas the visual display required two hands (one hand to hold the display and the other to use the input device). Second, the audio display does not require visual attention to perform the comprehension task. In our previous work examining different visual display technologies, participants noted difficulty in reading the passages because they kept "losing their line" in the text. By freeing the participants' visual resources they can be more fully used for processing the environment and paying attention to the path. More broadly, this might have beneficial implications for other mobile devices. Many users can be seen wandering the streets and corridors of offices with their head down staring and mobile email clients and web browsers. With an audio display option, they could still absorb and understand the text they are engaged in but regain the use of their eyes hopefully resulting in fewer accidental collisions with passers-by or objects in the environment.

# 7. FUTURE WORK

While our laboratory experiment examining reading comprehension while walking on a path provides interesting insights, we are also interested in exploring the capabilities of these displays in more natural settings and for use in more realistic tasks. Walking

the path involved navigating static obstacles, whereas a mobile device user in the real world would also encounter mobile obstacles such as other people. We are interested in assessing performance in navigating dynamic environments while using the audio display.

It is important to note that the laboratory environment provided a quiet environment for listening to the audio. While most everyday environments do have noise, a pair of high–quality noise–blocking headphones, which are commonly used with mobile audio players such as the Apple iPod, would lead to similar low–noise conditions in the everyday environment. Future work will involve assessing the effectiveness of the audio display in a more realistic audio environment (i.e. ambient noise).

As noted above, our experiment used comprehension trials designed to help students practice for standardized tests. We would also like to explore other comprehension tasks that are likely to be performed on mobile devices, such as browsing email or reading a web page. These tasks may see better performance in both of the display types, as the user has a better sense of background information and context. Additionally, in tasks such as email, users are more likely to have a personal interest in the material which may affect the user's ability to comprehend the material. One participant in our study said they would have done better (been able to get more information out of the passages) if they contained information they cared about.

Finally, while we only studied the use of the audio and visual displays separately for comprehending while walking, we are interested in how performance (for walking and for comprehending) would fare if information was provided through multiple channels at once. Would users be able to follow along and comprehend better if they both saw and heard the information at the same time?

## 8. CONCLUSIONS

We evaluated in–motion reading performance on mobile devices for both a handheld visual display and a speech–synthesis audio display. Overall, we found the audio interface allowed our participants to better navigate their environment. Furthermore, participants rated the audio interaction as less demanding than the visual display for "reading" while walking. Together, these findings indicate that users may benefit from an audio display. Having a speech synthesis display in mobile e–book readers, web browsers, and email clients would allow people to better use their mobile devices in more situations and on–the–go.

## 9. ACKNOWLEDGEMENTS

This work is funded in part by the National Science Foundation and the National Institute on Disability and Rehabilitation Research. This material is based upon work supported by the National Science Foundation (NSF) under Grant No. 0093291. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF. This is a publication of the Rehabilitation Engineering Research Center on Mobile Wireless Technologies for Persons with Disabilities, which is funded by the National Institute on Disability and Rehabilitation Research of the U.S. Department of Education under grant number H133E010804. The opinions contained in this publication are those of the grantee and do not necessarily reflect those of the U.S. Department of Education.

## 10. REFERENCES

[1] B. Arons. Speechskimmer: Interactively skimming recorded speech. In *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 187–196, 1993.

[2] S. Baker, H. Green, B. Einhorn, M. Ihlwan, A. Reinhardt, J. Greene, and C. Edwards. Big bang! BusinessWeek, June 2004.

[3] L. Barnard, J. S. Yi, J. A. Jacko, and A. Sears. An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. *International Journal of Human-Computer Studies*, 62(4):487–520, 2005.

[4] L. Barnard, J. S. Yi, J. A. Jacko, and A. Sears. A new perspective on mobile device evaluation methods (in-press). *To appear in Personal and Ubiquitous Computing*, 2005.

[5] S. Card, T. P. Moran, and A. Newell. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum, 1983.

[6] S. G. Hart and L. E. Staveland. *Human Mental Workload*, chapter Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. North-Holland, 1988.

[7] S. G. Hill, H. P. Iavecchia, J. C. Byers, A. C. Bittner, A. L. Zaklad, and R. E. Christ. Comparison of four subjective workload rating scales. *Human Factors*, 34(4):429–439, August 1992.

[8] M. Johnston, S. Bangalore, G. Vasireddy, A. Stent, P. Ehlen, M. Walker, S. Whittaker, and P. Maloor. Match: An architecture for multimodal dialogue systems, 2002.

[9] J. Lai, K. Cheng, P. Green, and O. Tsimhoni. On the road and on the web?: comprehension of synthetic and human speech while driving. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 206–212, New York, NY, USA, 2001. ACM Press.

[10] J. Lai, D. Wood, and M. Considine. The effect of task conditions on the comprehensibility of synthetic speech. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 321–328, New York, NY, USA, 2000. ACM Press.

[11] R. S. Laramee and C. Ware. Rivalry and interference with a head–mounted display. *Transactions on Computer–Human Interaction*, 9(3):238–251, September 2002.

[12] LearningExpress, editor. *501 Reading Comprehension Questions*. Learning Express, 1999.

[13] K. Lyons. Everyday wearable computer use: A case study of an expert user. In *Proceedings of Mobile HCI*, pages 61–75, 2003.

[14] Mobile CommerceNet http://www.mobile.seitti.com, January 2002.

[15] T. Mustonen, M. Olkkonen, and J. Hakkinen. Examining mobile phone text legibility while walking. In *CHI '04 extended abstracts on Human factors in computing systems*, pages 1243–1246, New York, NY, USA, 2004. ACM Press.

[16] D. K. Roy and C. Schmandt. Newscomm: A hand-held interface for interactive access to structured audio. In *CHI*, pages 173–180, 1996.

[17] N. Sawhney and C. Schmandt. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human Interaction*, 7(3):353–383, 2000.

[18] K. Vadas, K. Lyons, D. Ashbrook, J. S. Yi, T. Starner, and J. Jacko. Reading on the go: An evaluation of three mobile display technologies. Technical Report GIT-GVU-06-09, GVU Center, Georgia Institute of Technology, 2005.