Enhancing Mobile Reading with Automatic Mode Switching on Reading and Listening

Chen-Hsiang Yu and Robert C. Miller
MIT CSAIL
32 Vassar Street
Cambridge, MA 02139, USA
{chyu, rcm}@mit.edu

ABSTRACT
Although the web browser has become a standard interface for information access, mobile web browsing remains a challenge. This research focuses on enhancing mobile reading for sighted users. In this paper, we propose an approach to combine reading and listening with automatic and manual switching to support continuous mobile reading, which happens with frequent context switching, such as from sitting to walking or from walking to sitting. A systematic evaluation of 250 unseen samples on One-Against-All SVM classifiers shows that it is feasible to predict 2 defined states with 99.6% accuracy and 5 defined conditions with 93.2% accuracy. These states and conditions are used for automatic reading and listening mode switching. A user study with 10 mobile users shows that automatic mode switching not only provides the easiest reading experience for mobile reading, but it also provides significantly less dangerous encounters compared with purely visual reading on a mobile device. The *Read4Me* Browser is a prototype system built to demonstrate part of the evaluated ideas.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Design, Measurement

Keywords: Read4Me, mobile reading, mobile browsing

INTRODUCTION

According to Pew Internet Mobile Access 2010 [1], eight in ten American adults (82%) currently own a cellphone of some kind and 32% of American adults use their cellphone to access the Internet, an increase from the 25% in 2009. Although the mobile phone penetration has been increased in these few years, reading web pages on a mobile device is still not as easy and comfortable as doing the same activity on a desktop computer.

The mobile device is used with a short attention span and users are prone to switch activities while using the device. However, the mobile property of a mobile device raises another challenge, which was less addressed in the past -- how do we support continuous reading on a mobile device in the presence of frequent context switches, such as switching from sitting to walking or from walking to driving or cooking. Difference contexts impose different constraints, e.g., cooking and driving may restrict the user from holding the device.

Whereas focusing on the display of a mobile phone may be difficult while walking or driving, Vadas's study showed that audio is an acceptable modality for mobile comprehension of text. [17] Although audio has been used in screen readers for blind or visually impaired people [4,5,7,10], there was no clear suggestion of how to combine the audio modality with visual presentation for sighted people on mobile devices. We address this need in the domain of frequent context switching among mobile users and consider a research question: how do we help mobile users have a continuous reading experience when they switch contexts?

In this paper, we propose combining visual and audio reading modalities with automatic and manual switching approaches to enhance mobile reading for sighted people. These enhancements are meant to support mobile users who frequently switch contexts, and desire a continuous reading experience during these transitions. A user study with 10 mobile users shows that automatic mode switching not only provides the easiest reading experience for mobile reading, but it also provides significantly less dangerous encounters compared with purely visual reading in the presence of frequent context switching while reading on a mobile device. The *Read4Me* Browser is a prototype system built on the Android 2.3 platform to demonstrate part of the evaluated ideas.

The contributions of this paper are: (1) We propose the idea of combining visual and audio reading modalities with automatic mode switching to provide continuous reading on a mobile device. (2) We present a user study with 10 mobile users; the results show that automatic mode switching not only provides the easiest reading experience for mobile reading, but also provides significantly less dangerous encounters compared with purely visual reading on a mobile device.

In the rest of this paper, we will discuss related work and introduce the Read4Me browser from system design principles to implementation details. A systematic evaluation of mode switching detection and a user study of mobile reading will be presented. We will summarize some of the lessons learned from the user study in a discussion section before presenting our conclusions and plans for future work.

RELATED WORK

Visual Reading on Mobile Devices

Reading is the most common activity humans perform on PCs and mobile devices. Previous research shows that the difference between PCs and mobile devices affects the way users navigate websites. A usability study of viewing web pages on a mobile device presents the difficulty of reading web pages on a small display. Although research has shown that people use web services while moving from place to place, mobile users have a short attention span on a mobile device due to frequent context switches [12].

Audio Listening on Mobile Devices

Reading a document or a web page with audio listening on a mobile device is not new, but researchers are still investigating how to improve audio quality. Text-to-Speech (TTS) is the most common technique used to support audio listening, and its support on mobile device is increasing. For example, starting from Android 1.6, the Android platform provides a Text-to-Speech (TTS) feature which can be used in mobile applications for different purposes, such as translation, document and screen reading, etc. The iPhone SDK does not support TTS natively, but different third party APIs are available, such as iSpeech and Flite. Vadas's study indicates that audio is an acceptable modality for mobile comprehension of text. [17] However, there is no clear suggestion of how to combine the audio modality with visual presentations for sighted people on mobile devices.

Screen Reader on PCs and Mobile Devices

For visually impaired or blind people on a PC, a screen reader is a solution to access information on the Web. Screen readers include WebAnywhere [4], aiBrowser for Multimedia [10] and HearSay. On mobile devices, screen readers are a solution for sighted people to access information in a hands-free contex, such as cooking or driving. However, the screen readers on mobile devices, such as VoiceOver and Mobile Speak, still have relatively poor speech quality. Although Odiogo [11] is a web service that transforms blog posts and sites into near human quality audio speech, it cannot be used in real time inside mobile applications. The SeeReader is a mobile document reader, which combines TTS with automatic content recognition and presentation control, to notify users of important visual content while listening to the content. [Carter'09] However, it does not focus on supporting continuous mobile reading between frequent context switches.

Mobile Device Position Detection

Accelerometers have been studied and used for more than a decade. A number of interaction techniques have been explored to provide better reading experiences, such as tilt and flick for scrolling [6], physical orientation for exploring music [3], gesture recognition [8,13], activity recognition [2,9] and foot gestures [Scoot]. In this paper, we apply mobile sensor data to another application: continuous reading on a mobile device.

READ4ME BROWSER

Design Principle

Reading a web page with a long article on mobile web browser can be challenges, but frequent context switching, like standing up or getting in a car, make it even worse. This paper proposes using orientation sensor information to detect context switches and text-to-speech (TTS) to read web pages aloud, i.e., the browser can automatically detect certain conditions and react appropriately so the user can continue reading, even while switching between modalities that restrict use of the user's eyes or hands.

Read4Me Browser defines two operational states, Look & Point and Hear & Say, and uses orientation information to provide seamless transitions between these conditions. While the first state represents the condition of looking and pointing at a mobile screen, the second state supports hands-free conditions, such as walking or driving. In the Look & Point state, a user can read web pages as usual and use their finger to touch the screen for scrolling, clicking, zooming, etc. On the other hand, when switching to the Hear & Say state, users can listen to web page content, which is parsed by Read4Me Browser and read using TTS. In the Hear & Say state, Read4Me Browser reads one sentence at a time and highlights the sentence being read in red so that the user can easily identify it when the user switches back to the Look & Point state.

There are five conditions defined in Read4Me browser: (1) left-ear listening (2) right-ear listening (3) in docking (4) in pocket and (5) static reading. The first four conditions can trigger switching from the Look & Point state to the Hear & Say state. On the other hand, when the user places the phone in a viewable position, such as a (5), the phone will switch to the Look & Point state.

In addition to automatic mode switching, we also learned from the user study that it is also useful to include the ability to switch states manually. We took this suggestion into consideration and created another reading support feature, manual switching, which allows mobile users to manually switch between the Look & Point state (Reading) and Hear & Say state (Listening). This feature is useful in some situations. For example, when a mobile user is working in an office and feels tired of reading on the screen, they can close their eyes and enjoy listening to the content being read by the Read4Me browser.

System Overview and Implementation

The Read4Me Browser is designed for sighted people and keeps all the mentioned design principles in mind. To experiment with these ideas, we developed the Read4Me browser using the WebKit engine with the WebView widget on the Android 2.3 platform. Although we hope to develop a general mobile browser to be used on different platforms, cross platform compatibility has not been a priority at this stage.

User Interface Design

In order to make it easy to use for mobile users, Read4Me browser is designed to have a similar user interface as a normal mobile web browser. It has a URL bar, select button, content area and menu. The menu provides six options: Auto Switch (for automatic mode switching to support mobile reading), Manual Switch (for manual switching between reading and listening), Speech Recognition, Reload, Back and Forward. (Figure 1)

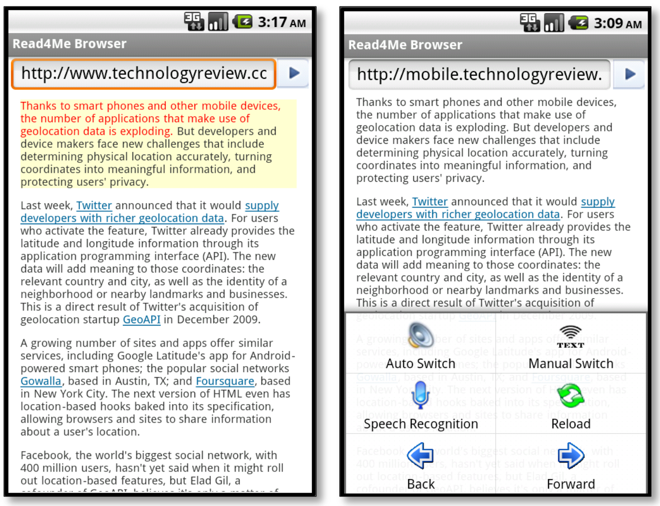

Figure 1: The Read4Me browser reading interface

MODE SWITCHING DETECTION FOR MOBILE READING

In order to detect mode switching on an Android phone, we use the orientation sensor to detect five postures: left-ear, right-ear, in-dock, in-pocket and static-reading. The coordinate system the phone has three axes: the X-axis is horizontal and points to right of the phone, the Y-axis is vertical in the plane of the phone's screen, and the Z-axis points directly out of the screen. The orientation sensor data can be represented by a set of (azimuth, pitch, roll) values as follows:

- Azimuth: the angle between the Y-axis and the magnetic north direction around the Z-axis (0-359 degrees)
- Pitch: the angle around X-axis (-180-180 degrees)
- Roll: the angle around Y-axis (-90-90 degrees)

This research collects orientation sensor data from HTC NexusOne smartphone, which is an Android 2.3 platform, to train and evaluate mode switching. We collected 250 samples from each of the five postures, 1250 samples in total. The samples were collected in a laboratory setting by the first author. Each sample contains azimuth, pitch, and roll values for a given posture label. The whole dataset is separated into a training dataset (1000 samples, 200 for each condition) and a testing dataset (250 samples, 50 for each condition) for the evaluation.

Training SVM Classifiers

We use the azimuth, pitch, and roll data from the orientation sensor to identify the state which mobile web browser should switch to for providing listening or reading support.

We trained Support Vector Machines (SVM), using one-against-all classification with a radial basis kernel, to predict the five defined conditions. To make each dimension have the same dynamic range, we divided the azimuth value by 200, the pitch value by 100, and the roll value by 50. In addition, we set the slack penalty to 10 and maximum number of iterations of quadratic programming to 100,000 during the SVM training.

To evaluate the five trained SVM Classifiers for five conditions, we use the testing dataset. We evaluate the performance of the classifiers in two ways. First, we looked at the accuracy of predicting the five postures. Second, we looked at prediction accuracy of the states triggered by those conditions: i.e. Look&Point State (which corresponded to the static-reading posture) and Hear&Say State (which corresponded to the other four postures). The results of the evaluation on testing datasets are shown in Table 1.

Evaluation Dataset: Testing Dataset (250 samples)

Prediction	# of samples	# of errors	Success Rate
2 States	250	1	99.6%
5 Conditions	250	17	93.2%

Table 1: Evaluating testing dataset on trained SVMs.

The results suggest that one-against-all SVM classifiers can accurately predict the appropriate state from orientation data.

USER STUDY ON MOBILE READING

Introduction

To understand the effect of the combination of reading and listening on mobile reading, we conducted a user study with 10 participants. The study focuses on reading web pages containing several paragraphs, forcing pay more attention to understand the content. The study is modeled on Vadas *et al.* study [17], which examined overall performance and perceived workload for four conditions: (1) audio-walking, (2) audio-sitting, (3) visual-walking, and (4) visual-sitting. There are three differences between our study and Vadas *et al*. First, we examined four different conditions: (1) Only visual reading (2) Only audio listening (3) Manually switching visual reading and audio listening (4) Automatically switching visual reading and audio listening. Second, we interrupt participants every minute to force them switch condition frequently from sitting to walking or from walking to sitting. Third, we use real web pages as reading material, which are lengthy and condense and need more user's attention in reading, but Vadas's uses a short passage, which only contains one to three paragraphs, in each trial and 5 trials for each condition. The third difference is that we ask participants to answer the questionnaire in paper format, whereas Vadas's uses mobile devices.

Procedures

The user study environment setting is illustrated in Figure 3. The path is taped on the ground and is 11.81 inches (30 cm) wide and 139 feet 5 inches long. All participants are informed to walk inside the taped path, but we recorded the number of out of path steps during the study.

There are four web pages: two articles selected from technology websites and another two from business websites. The order of web pages presented is randomly selected as well as the condition applied to those pages. Before the formal study, the experimenter presents one sample web page and allows the subject to play with the system to get used to reading web pages with automatic context switching mode. After each web page reading, we present a questionnaire, which contains 5 single choice comprehension questions and 2 subjective questions. After each study, we conducted a post-test interview to solicit subjects' subjective feedback.

During the testing, the experimenter raises a big sign and rings a bell to inform participants to switch activities, e.g, from walking to or vice versa. The switch happens every minute. For all conditions, all participants wore a Bluetooth headset and were asked to hold the phone as they walked through the path. After the study, we asked all participants to sit and answer a paper.

There are four conditions investigated in this study: C_1: Using only visual reading, C_2: Using only audio listening, C_3: Manually switching visual reading and audio listening, C_4: Automatically switching visual reading and audio listening. The order of reading articles and applied methods are pre-randomly arranged to balance each condition.

Figure 2: Holding a mobile device and walking inside the path.

Conditions C_2, C_3 and C_4 highlight the current listening sentence in red and the current listening paragraph with a yellow background (Figure 4). However, there are two differences between C_3 and C_4. The first is that one needs to manually press a button to enable or disable listening to a web page. The second is that when switching to visual reading state, manual switching removes the mark added on the current reading sentence. However, the listening paragraph is always shown at the beginning of the screen. While we intended C_3 to keep the mark across mode switches, a bug in the browser prevented the highlighting from displaying. Inside the mobile web browser, we provide two buttons for condition 3 and 4, i.e. manual switching or automatic switching.

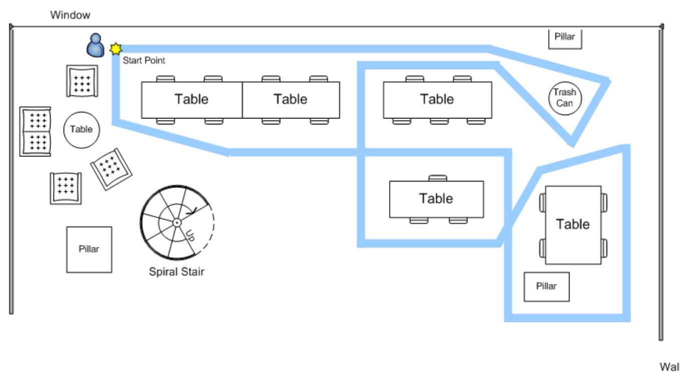

Figure 3: The floor print of user study experiment. (The blue line is the path for participants to walk through.)

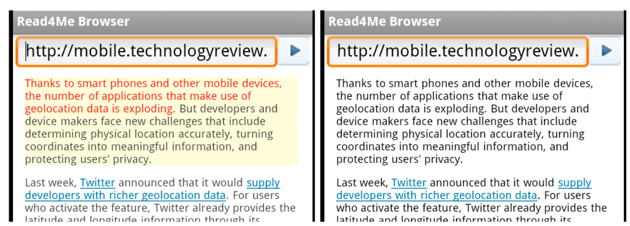

Figure 4: The difference between marked sentences and paragraphs (left) with plain ones (right).

In this user study, we measure: (1) reading speed (2) reading comprehension (3) number of out of path steps (4) user satisfaction in ease of reading (5) user satisfaction in self-evaluation of understanding.

Hypotheses

The user study was designed to address the following hypotheses:

(H1) Using only visual reading provides the best reading comprehension, but it also has the highest out of path rate.

(H2) Using only audio listening provides the best user satisfaction.

(H3) Automatic mode switching does not offer similar reading comprehension as using only visual reading, but automatic mode switching provides the best user satisfaction among all four methods.

(H4) Manual mode switching provides a similar reading comprehension as automatic mode switching, and it has a lower user satisfaction compared with automatic mode switching.

Participants

We used a within-subjects study design. We recruited ten participants using emails and flyers at our university. There were 8 males and 2 females, on average 26.6 years old. Seven of the participants had not listened to audio books in the past. All of them had used mobile web browsers, and all were fluent in English.

Results

We measured the reading speed, reading comprehension, and number of out-of-path steps. We supplemented these results with subjective measures of user satisfaction with each reading approach and users' confidence that they understood the content. We analyzed the results using a one-way ANOVA.

Figure 5 presents the average reading speed for each of the four conditions, measured in words per second. There is no significant difference in this measurement.

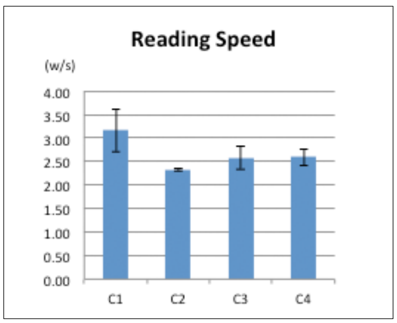

Figure 5: Reading speed for each condition, in words per second. Error bars show standard error.

Figure 6 shows the result of our reading comprehension measure, which is defined as the number of correct answers out of 5 comprehension questions. There is no significant difference in this measurement.

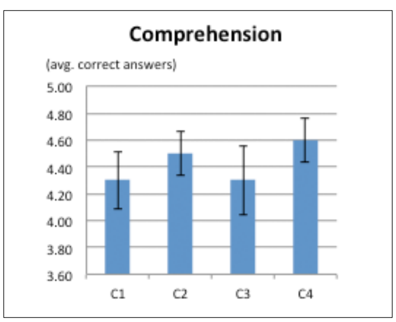

Figure 6: Comprehension for each condition (measured by number of correct answers to a 5-question quiz).

Figure 7 presents the number times the user stepped out of the taped path in each condition. The experimenter kept track of this value as the participant walked, counting a step as out-of-path if half of the foot area was outside the blue taped path. There is a significant difference in this measurement. ($F_{3,36} = 3.53$, $p < 0.05$) A post-hoc Tukey test reports an honest significant difference between C_1 and C_4 on the number of out-of-path steps and also gives a borderline honest significant difference between C_1 and C_2 ($p < 0.05$).

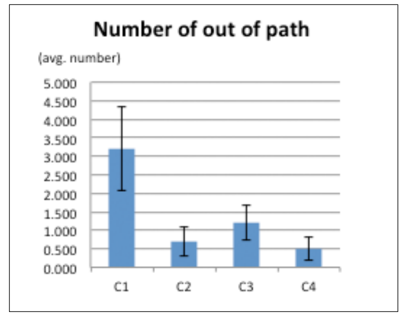

Figure 7: The number of out of path for each condition. (measured by the number of out of path).

Figure 8 shows subjective reading ease of the web page, on a Likert scale (1 = very difficult, 7 = very easy). There was a significant difference in subjective reading ease ($F_{3,36} = 8.88$, $p < 0.001$). A post-hoc Tukey test gives an honest significant difference between C_1 and C_4 ($p < 0.001$).

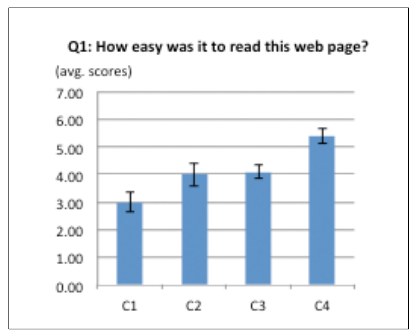

Figure 8: Ease of reading (rated on a scale from 1=very difficult to 7=very easy).

Figure 9 reports on the second subjective question, self-reported comprehension. Self-reported comprehension was not significantly different between conditions.

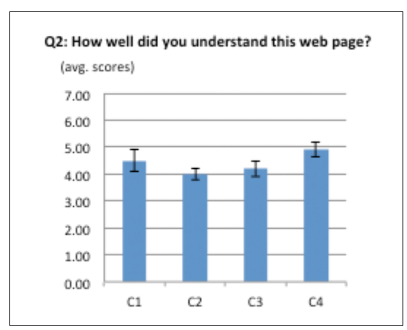

Figure 9: User's self-evaluation of understanding (1=very difficult, 7=very easy).

Based on the results we learned from the user study, we can answer part of our hypotheses:

(1) Under the condition of frequently switching between walking and sitting and doing mobile reading, using only visual reading on mobile device has a similar reading comprehension as listening on mobile reading, manual and auto

switch listening and reading on mobile device. However, it does have the highest rate of stepping out of path, which might be a danger to mobile users.

(2) Using only audio listening does not provide the best user satisfaction. Automatically switching between listening and reading on the mobile browser provides the best user satisfaction.

(3) Automatic mode switching not only offers a similar reading comprehension as using only visual reading, but it also provides the best user satisfaction among all four methods. In fact, there is a significant difference between using these two methods in mobile reading.

(4) Manual mode switching provides a similar reading comprehension as automatic mode switching, but there is no significant difference in user satisfaction between manual and automatic mode switching.

DISCUSSION

There are many useful comments we received from the participants and we can categorize them into four categories:

Visual reading is not always good

Reading on the go is difficult, but many people still do it occasionally in their daily life. Some comments support this point: "(It is) Very difficult when walking, because I couldn't balance well and my eyes hurt" (User_1), "[I] can quickly scan document and automatically adjust pace to document location that require more thorough understanding" (User_5), "… it could be dangerous if not inside a room" (User_6). However, several users responded that they have better reading comprehension in visual reading than listening. "Comprehension is better with (visual) reading compared to pure listening" (User_3), "I can read much more quickly than something can be read. I can also skim over portions I know things about …" (User_8).

Listening isn't simply listening

Compared with reading on the go, users prefer listening on the go, which matches the findings in Vadas's work. A common suggestion we hear from users is that they need more control in the audio listening condition. Because of the potentially distracting background environment, such as surrounding people, it is easy for a mobile user to lose their focus on listening and be unable catch up in their understanding. As User_1 notes, "I prefer having control while listening (reading-like experience while listening, such as repeat, jump ahead, change speed, etc.)." User_3 adds, "The inability to go back while listening seems to be a fundamental problem with listening because it might be very hard for the system to guess exactly what I want to listen to again." The preferred way to trigger actions might be "voice command" or line control. Providing improved quality audio over TTS and support for quickly glancing at content while listening, e.g. to key sentences, are another two requests.

Mobile Web browser should be smarter

One issue that comes out from the user study concerning the system is that when switching from reading to listening, the browser is not smart enough to know where the user stopped reading. Instead, it normally reads the start of the first visible paragraph. User_2 mentioned, "Sometimes when I stopped reading in the middle of A, but the speaker start from B directly, and I lost some part of A (this causes me sometimes to go back and re-read something.)"

Manual and automatic switching modes are useful

While five out of ten users explicitly mentioned that automatic mode switching in mobile reading is their most-preferred method, three out of ten preferred manual mode switching. It would be helpful to provide both of these two features in the mobile web browser to help mobile reading. One suggestion from users is that we should also provide in manual mode the same visual cues used in automatic switching mode, so that they can easily understand where they stopped listening when they switch to reading.

There are two challenges in using TTS to support the audio-walking condition. First, content filtering is necessary to enhance the listening comprehension. There are three types of information might be a distraction. (1) Abbreviation: for example, in reading-sitting or reading-walking condition, reading TX or CO in the context can be interpreted as Texas or Colorado intuitively. However, in the audio-sitting or audio-walking situation, speaking TX or CO does not carry the same meaning as speaking Texas or Colorado. (2) Punctuation: in addition, filtering punctuation is also necessary. Without transforming the content, "--" will be pronounced as "hyphen hyphen". This disrupts the reader's listening and annoys the reader. Instead, speaking with a short pause is acceptable to handle such text. (3) Foreign names or place names: uncommon foreign names or location names might be fine in visual reading, but will cause difficulty for listeners, such as with Bhattachariee, Chávez, etc.

CONCLUSION AND FUTURE WORK

Mobile reading and browsing has many limitations originating from its physical constraints. Our research attempts to enhance mobile reading for sighted users under the condition that they might have a frequent context switching, such as switching from sitting to walking or from walking to driving or cooking. The main challenge in this scenario is that context switching will interrupt the original reading process and the user might not understand where to start again, and may well just give up.

In this paper, we proposed to combine the visual reading and audio listening with automatic and manual mode switching to provide continuous reading on mobile devices. To exemplify the automatic mode switching detection, we use One-Against-All SVM with Radial Basis Kernel to predict the five defined conditions. The evaluation on five trained SVM classifiers shows that it can predict five predict the five defined conditions. The evaluation on five trained SVM classifiers shows that it can predict five predefined conditions with 93.2% accuracy. It can also discern between two predefined states, Look&Point and Hear&Say, with 99.6% accuracy. The results of the user study with 10 mobile users shows that using automatic mode switching provides the best mobile reading experience in situations with frequent context switches. There is a significant difference between using this approach with purely using visual reading in the same condition. In addition, Automatic mode switching also results in significantly less steps out of the path, which might avoid dangerous situations.

Read4Me browser is a prototype system built on top of the Android platform to demonstrate some of the evaluated ideas. Read4Me browser provides both automatic mode switching and manual mode switching. The browser demonstrates the feasibility of a speech command interface for interacting with mobile web browsers in reading tasks. We will continue to design features to support audio listening, such as forward, rewind, speak faster, speak slower, skim this paragraph or jump to next paragraph, voice and emotion annotation, etc. Another interesting direction for exploration is how to allow mobile users to share the mobile reading experience with his social networks. For example, mobile user can annotate his current page such that the friends in his social network can read the annotations when they read the same web page. We look forward to studying the effect and impact of such mutual interactions.

ACKNOWLEDGMENTS

REFERENCES

1. Aaron Smith. Mobile Access 2010, Pew Internet & American Life Project, 2010.
2. Bao, L. and Intille, S.S. Activity recognition from userannotated acceleration data. In Proc. Pervasive 2004, Springer (2004), 1-17.
3. Bergman, J. et al. Hands on music: physical approach to interaction with digital music. MobileHCI, 2009.
4. Bigham, J., Prince, C. and Ladner, R. Webanywhere: A Screenreader On-the-go. *W4A '08*, 2008.
5. Freire, A. et. al. Using screen readers to reinforce web accessibility education, *Proceedings of the 12th annual SIGCSE conference on Innovation and technology in computer science education*, 2007.
6. Fitchett, S. and Cockburn, A. Evaluating reading and analysis tasks on mobile devices: a case study of tilt and flick scrolling, Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7, 2009.
7. Gaal, A. et. al. Towards an Open Source Screen Reader: Screenreader Usability Extensions (SUE). *In Proceedings of ICCHP*, 2008.
8. Hudson S. E., Harrison, C., Harrison, B. L., and LaMarca, A. Whack gestures: inexact and inattentive interaction with mobile devices. In Proc TEI 2010, ACM (2010), 109-112.
9. Iso, T., and Yamazaki, K. Gait analyzer based on a cellphone with a single three-axis accelerometer. In Proc. MobileHCI 2006, ACM (2006), 141-144
10. Miyashita, H. et. al. Aibrowser for multimedia: introducing multimedia content accessibility for visually impaired users. *Proceedings of SIGACCESS on Computers and accessibility*, 2007.
11. Odigo: http://www.odiogo.com/
12. Oulasvirta, A. et al. Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile HCI. In Proc. CHI 2005, ACM Press (2005), 919-928.
13. Patel, S.N., Pierce, J.S. and Abowd, G.D. A gesture-based authentication scheme for untrusted public terminals. In Proc. UIST 2004, ACM (2004), 157-160.
14. Ramakrishnan, I. et al. Hearsay: enabling audio browsing on hypertext content, Proceedings of the 13th international conference on World Wide Web, May 17-20, 2004, New York, NY, USA.
15. Sawhney, N. and Schmandt, C. (2000). Nomadic Radio: Speech and audio interaction for contextual messaging in nomadic environments. ACM Transactions on Computer-Human In- teraction, 7, 353–383.
16. Scott, J., Dearman, D., Yatani, K. and Truong, K.N. 2010. Sensing Foot Gestures from the Pocket. In *Proceedings of UIST 2010, 2010.*
17. Vadas, K., Patel, N., Lyons, K., Starner, T. and Jacko, J. Reading on-the-go: a comparison of audio and handheld displays. In MobileHCI, 2006.