Sasayaki: Voice Augmented Web Browsing Experience

Daisuke Sato Shaojian Zhu† Masatomo Kobayashi Hironobu Takagi Chieko Asakawa

IBM Research – Tokyo
1623-14 Shimo-tsuruma, Yamato, Kanagawa, Japan
{dsato, mstm, takagih, chie}@jp.ibm.com
+81 46 215 {4793, 4679, 4557, 4633}

†UMBC
Baltimore MD 21250, United States of America
szhu1@umbc.edu
+1 410 455 3883

ABSTRACT

Auditory user interfaces have great Web-access potential for billions of people with visual impairments, with limited literacy, who are driving, or who are otherwise unable to use a visual interface. However a sequential speech-based representation can only convey a limited amount of information. In addition, typical auditory user interfaces lose the visual cues such as text styles and page structures, and lack effective feedback about the current focus. To address these limitations, we created *Sasayaki* (from 'whisper' in Japanese), which augments the primary voice output with a secondary whisper of contextually relevant information, automatically or in response to user requests. It also offers new ways to jump to semantically meaningful locations. A prototype was implemented as a plug-in for an auditory Web browser. Our experimental results show that the *Sasayaki* can reduce the task completion times for finding elements in webpages and increase satisfaction and confidence.

Author Keywords

Voice Augmentation, Voice Browser, Accessibility.

ACM Classification Keywords

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces; K.4.2 [**Computers and Society**]: Social Issues – *Assistive Technologies for Persons with Disabilities.*

General Terms

Experimentation, Human Factors

INTRODUCTION

The Web represents one of the largest paradigm shifts in history. One out of every four people uses the Web [1] with great benefits. At the same time many people are unable to access the Web with standard visual browsers because they have visual or reading limitations, have only a basic and screenless mobile phone, or have constraints while moving or are otherwise unable to read a display. Given that an estimated one billion people have visual or reading limitations (est. 700 million with limited literacy [2] and 300 million with visual impairments [3]) and that about 4.6 billion mobile phones are in use [4], auditory interfaces could greatly expand access to the Web.

The current capabilities of auditory interfaces are quite limited and the Web is still rapidly evolving, making many webpages too complicated to access with vocal interfaces such as screen reading software. There is a central limitation in the current model of Web navigation. An auditory user interface is sequential, so it provides only a limited amount of information each time, which makes it hard for the users to get an overview of the webpage. Most webpages have highly visual layouts that are hard to understand by ear. Due to these limitations, people who use auditory interfaces to browse webpages find it slow and also tend to lose their places, making their browsing experience less reliable and less satisfying [5].

Therefore, we devised a new system called *Sasayaki* (which means 'whisper' in Japanese) to improve the Web experience of people who are using auditory interfaces rather than visual presentations. *Sasayaki* aims to simulate a friend of the user who is standing nearby with whispered tips for navigating the webpages. This virtual friend is watching the display and observing the behavior of the user to provide context-based hints and guidance. Computers cannot simulate complicated human behaviors, but even partial assistance can be helpful. Various kinds of information that cannot be produced by straightforward text-to-speech technologies can be provided by using heuristics and data analysis techniques with the data from the webpages and records of user behaviors.

As the first proof-of-concept prototype, the *Sasayaki* browser we developed presents the output of a standard auditory browser and the supporting information about the webpages in parallel through separate voice channels. The main voice provides raw text and some structured information from the webpages. In contrast, the secondary voice is for contextual navigation guidance based on the user's position and overviews of the webpages from content analysis and the metadata for the webpages. A user listens

to these two different voices from two physically separated speakers. *Sasayaki* helps users dynamically switch their focus depending on the context, so they can focus on relevant information while moving and jumping quickly, but without losing track of their location within the webpage.

This paper summarizes related work and describes the concept and a prototype of the *Sasayaki* along with a user study involving nine blind participants. Although the prototype and the initial evaluation focus on the auditory Web browser for people who are blind, the insights obtained will be helpful for developing broader types of auditory user interfaces.

RELATED WORK

Our work was initially inspired by observing the urgent needs of screen reader users to overcome problems as they were surfing the Web. Barnicle [6] evaluated screen reader users' experiences with thirteen visually impaired people and identified 58 unique obstacles that are not problematic to the same degree for sighted people. Lazar et al. [7] collected users' reports of the causes of frustration when using screen readers, with confusing page layout being ranked as the leading problem. They also found that a user generally spends more time recovering from an incident than the time it took to create the problem.

User Interface Agent

Sasayaki was also guided by the concepts of user interface agents, which unobtrusively provide users with needed help. Maes [8] talked about the concepts of interface agents to help users save work and avoid information overload. Among the various types of user agents, we focused on two categories.

Voice Augmentation

Bederson [9] created an automated tour guide prototype that uses audio to guide tourists. Sawhney and Schmandt [10] worked on 'Nomadic Radio', a wearable system which allows users to access information by using voices and textual information in a nomadic environment. This agent system could decide how to most effectively present information to the user based on the context, interruption settings, and automatic text understanding. Eckel [11] created the Listen project, which augments everyday environments with interactive soundscapes. Depending on the location and other context data, the system can suggest the most effective options for users. Other voice-based agent systems seek to provide users an impression of reality. Kalantari et al. [12] and Miyashita et al. [13] wrote about their voice-based augmented reality systems for visitors at museums. Voice-based agents were also used for navigation. Shoval et al. [14] introduced NAVBELT and GUIDECANE, which are voice and touch-based tools that can use a stereo earphone with a tactile stimulator to help blind people navigate in real environments. Jones et al. [15] described the ONTRACK system which uses adaptive music playback to support navigation in a 3D virtual space.

Browser Agents

Web surfing is one of the important tasks for screen reader users when they interact with computers. Stylos et al. [16] introduced an intelligent clipboard monitoring agent that helps to identify formatted data (such as addresses or appointments) for smart pasting into webforms. Wagner and Lieberman [17] introduced Woodstein, which predicts and assists the next user action based on analysis of a collected sequence of previous actions on the webpages. Roth et al. [18] created an agent to provide audio feedback for the user's cursor location. Yu et al. [19] created context-aware Web agents to provide audio and haptic feedback for the user's cursor location in the screen reader. Dontcheva et al. [20] created a Web agent that can help record and organize user sessions for comparison and analysis. The authors reduced the users' memory load and simplified tasks. Hartmann et al. [21] described Augur, a context-based smart agent that can do three things for users: highlight, suggest, and automate by analyzing context data and using pre-defined rules. Parente [22] introduced Clique, an auditory interface with four assistants and distinct voices. Each assistant has a role involving tasks or events on a desktop including email, calendar, and browser applications. Although sometimes the assistants speak simultaneously, it is a different type of synchronicity from *Sasayaki* concepts.

Context Awareness

Context has also received attention for more intelligently improving accessibility. Most user agents systems [10, 14] make good use of context for decisions. Mahmud et al. [23] introduced their CSurf system that finds the most relevant information based on user behaviors and other contextual clues using a statistical machine-learning model. Borodin et al. [24] reported on the problem-causing lack of awareness of visual changes in dynamic webpages. This led to an algorithm to detect changes in dynamic webpages, allowing the system to collect useful context information. As in Web browsing scenarios, they found that user behavior, cursor location, and webpage layout were the most important contextual clues for analysis [17, 19, 21, 23].

Assistive Technologies

Sasayaki was guided by the previous work on adding intelligence to screen readers. Various approaches have been tried to help screen reader users have better browsing experiences by adding more powerful functions to the software. Yesilada et al. [25] introduced DANTE, a semantic transcoding system for webpages visited by blind people. DANTE can transform a typical webpage into a screen-reader-friendly format. Harper and Patel [26] described the effects when visually impaired people are unable to scan webpages. They created a system to summarize webpages for blind Web surfers. Miyashita et al. [27] created a multimedia voice browser that handles the multimedia objects on webpages by adding scripts to give visually impaired users full control over the objects. Lunn et al. [28] introduced SADIe, which uses semantic

annotations in a CSS (Cascading Style Sheet) to transform webpages for better accessibility in a screen reader.

There are also studies focusing on employing the power of social networking to add smart tagging information to existing webpages to improve their accessibility. Takagi et al. [29] created Social Accessibility to enable third-party volunteers to effectively assign layout tags to webpages. The screen reader users have special tools to retrieve and use this metadata to improve their surfing experience. Chen and Ramen [30] created AxsJAX, which lets programmers dynamically insert ARIA statements into webpage content so that screen reader users can handle dynamic pages.

These technologies try to improve Web accessibility by transcoding content and also to provide better experience for the users based on the current single voice browsing interface. *Sasayaki* can coexist without conflict with these technologies and can further improve the Web experience by working with them.

SASAYAKI

We regard *Sasayaki* as an example of a general concept of a user interface that provides supplemental information for many kinds of people via an audio channel. From the users' various viewpoints, ideally they would have a non-intrusive audio-based augmented environment. *Sasayaki* tries to act in a supportive role to help the users as would a sighted person in assisting the visually impaired user (Figure 1). It provides apparently intelligent feedback by analyzing user behavior, status, and other related contextual evidence. *Sasayaki* then provides bits of verbal information and advice with a synthesized voice. *Sasayaki* can also provide other kinds of audio-based hints, such as sound effects, background music, or background noises. One simple example is that a car navigation system could provide extra information about popular restaurants near the anticipated location of the user's car based on the planned route, in parallel with the main voice that provides driving directions. Another example would help a blind user shopping for gifts on an e-commerce site. In this situation, the agent could suggest popular gifts while helping to orient and guide the user around the webpages. Given that people who have visual impairments are supposed to be the main population who must depend on the auditory interface and often experience various difficulties in using it, we focused on them in the initial design and development of *Sasayaki*.

Categories of Sasayaki Information

The following four categories of information are expected to be useful for blind users navigating webpages. This includes information that is hard to obtain from non-visual user interfaces as well as some information that is useful for sighted people.

Spatial: For the blind users, it is very hard to be aware of the position of their cursors or to know the position where the screen reader is currently reading. Although this is very fundamental and easy for sighted people, the blind users cannot get an overview of their context without special help. There are Braille blocks, Braille maps, and audio signals for crossing streets or climbing stairs in our physical world, and these give blind people some awareness of their locations and the context. Webpages need similar signals for easier navigation [31].

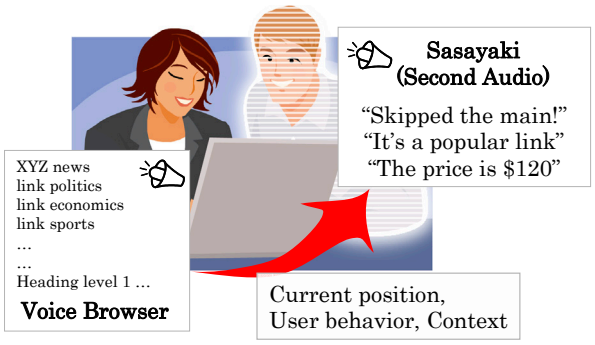

Figure 1. Concept of Sasayaki User Interface Agent

Contextual: *Sasayaki* changes behavior according to the context of the users. For example, the *Sasayaki* system can detect when a user seems to be lost and then provide appropriate support to recover. As described in the related work section, there are many forms of contextual support. For example, when a user is shopping, the *Sasayaki* provides prices and user reviews for achieving the goals of the user's tasks.

Social: *Sasayaki* provides a kind of extra intelligence for blind users. The intelligence can be socially generated by volunteers or users. Many studies tried to improve accessibility by using external metadata [27, 28, 29, 30]. Such metadata can provide an outline of a webpage, locating a header, the main content, advertisements, related information, and a footer.

Analysis: The intelligence also can be generated by the computer automatically. Analyzing many user reviews to summarize the users' positive and negative reactions to a restaurant is a typical example. Another example from an e-Commerce page would be to suggest popular links or buttons in webpages by analyzing clickstreams, which would allow them to be aware of which webpages other people tend to navigate to without any visual cues..

IMPLEMENTATION OF PROTOTYPE SYSTEM

We have implemented a prototype of the *Sasayaki* system. Our prototype helps users to retain awareness of their current position while providing overviews of the webpages. This supportive information becomes available to the users through the whispering channel, either automatically or as requested. *Sasayaki* comes from a different sound device physically separated from the main speaker and uses a different speech synthesizer engine to simulate the "whispering" scenario. In this way, *Sasayaki* efficiently and simultaneously presents both voices to the users, making

Sasayaki less intrusive but still noticeable. The *Sasayaki* system is designed as a plug-in component for our voice browser called aiBrowser [32]. The behaviors and functions of aiBrowser are similar to popular screen readers. Since *Sasayaki* system needs an API to obtain the context of the Web browsing, aiBrowser was extended to provide it. *Sasayaki* could work with other screen readers by adding a corresponding API as a plug-in or extension.

Role-based Notifications and Jumps

To provide users with contextual support, the *Sasayaki* plug-in retrieves pre-defined role-based data about a webpage from an Accessibility Commons [33] server. The typical roles for content include main content, header, advertisement, and so on. The *Sasayaki* plug-in also monitors the position of the current focus of the voice browser and tracks key events linked to user behaviors. This allows the plug-in to generate the most appropriate advice for each user. For example, when a user reaches the defined main content body for the first time, the system will decide that "Entering main" is a good contextual prompt. We call this *Sasayaki* whispering. It may also include more advanced help for the users such as "Skipped the main content". The *Sasayaki* whispering is provided at several role-tagged parts of the webpage. *Sasayaki* also allows users to freely change the focus of the screen reader between role-based parts using a 'jump' function. We decided to add this feature as jump functions are very common in modern screen reading software. For example, many users already use jumps between heading elements (H1-H6), table elements, list items, form elements, and so on. Hence users should benefit from the *Sasayaki* jump functions without changing their interaction model for browsing webpages.

Page Overview and Text Analysis

In addition, for webpages that belong to the same category, the important information tends to be similar. For example, for product pages from an online shopping website, each page has the same kinds of important information, including price, in-stock information, shipping rules, and so on. On complex webpages, non-visual users may have difficulties in simply finding that basic information. *Sasayaki* can collect the basic information for the page category and output it in an efficient format as requested by using metadata, so that the users can quickly get an overview of the crucial content in each page.

Another problem is when voice browser users want the useful information in a large body of user reviews. Some online shopping webpages collect such reviews, and in this case *Sasayaki* can extract and analyze a large number of user reviews by using its sentiment-based text mining component [34]. *Sasayaki* does this by retrieving the original user reviews as text from the related webpages and passing it internally to the text-mining component that generates statistics about the user comments, covering various product aspects such as price, quality, or texture.

EVALUATION

An empirical study was conducted to evaluate the user performance and behaviors in navigating webpages using our *Sasayaki*, comparing it to typical screen reader software. The primary foci of the evaluations were the role-based notifications and jump functions. Each participant was familiarized with *Sasayaki* and then asked to perform 5 tasks. After observing their performance, we also asked for their subjective ratings of the agent system.

Pilot Study

A pilot study [35] was done before the main experiment with three blind people (2 males and 1 female from 37 to 44 years old). Two of them are completely blind and one has limited vision. Four tasks and a survey were given to each participant. In this pilot study, we tried to explore differences between the same system with and without *Sasayaki* functions. We compared the original aiBrowser system and aiBrowser with all of the *Sasayaki* functions. We found that *Sasayaki* effectively supported the users' navigation. The users working with *Sasayaki* spent much less time navigating to the required page elements for each task. They also showed high confidence in their abilities to do Web browsing with *Sasayaki* functions. This study supported our belief that *Sasayaki* could be a useful aid to improve user experience. This led us to focus on the differences between *Sasayaki* whispering functions and the *Sasayaki* jump functions in the main experiment.

User Experiment

Nine native Japanese blind people (8 males and 1 female, from 30 to 53 years old), none of whom were involved in the pilot study, participated in the experiment as paid volunteers. They are referred to as P1 to P9 in the following sections. All of them were completely blind and all had experience with Web navigation using screen reading software. Eight of the participants were expert users with Web experience going back to the '90s. The participant P9 became blind a few years ago and also has less experience with the Web. Most of the participants had little or no exposure to the webpages used in this experiment, though one participant was already familiar with the Amazon.co.jp site. None of them had experience with aiBrowser.

Equipment

A ThinkPad T400 laptop with a 2.40-GHz Core 2 Duo CPU running Windows XP was used as the experimental computer. This laptop has a build-in stereo speaker which was used for the main voice output. A Yamaha USB-powered stereo speaker (NX-U10) was placed next to the laptop for the *Sasayaki* voice. A standard Japanese 109-key USB keyboard was used instead of the laptop keyboard, because keyboards differ among laptops and using a standard Japanese keyboard helped avoid confusion.

Sasayaki Conditions

We compared two types of differences, whether or not the system has the *Sasayaki* whispering function and whether

or not the system has the *Sasayaki* jump function. There were four test conditions for the *Sasayaki* settings to evaluate how the *Sasayaki* whispering and the *Sasayaki* jump functions affected the participants' performance of typical navigation tasks.

The first condition was the original aiBrowser (**NS-NJ**). The second was aiBrowser with *Sasayaki* whispering but without *Sasayaki* jump (**S-NJ**). The third was aiBrowser with the *Sasayaki* jump function but without *Sasayaki* whispering (**NS-J**). The fourth was aiBrowser with both *Sasayaki* whispering and *Sasayaki* jump (**S-J**).

Tasks

We used a within-participant design, so for each webpage each participant was asked to perform the same set of tasks within each of the four *Sasayaki* conditions. The independent variables were whispering (with or without *Sasayaki* whispering) and jump (with or without *Sasayaki* jump). There were four sets of five target webpages (news article on Asahi[1] and Nikkei[2], product page on Amazon[3] and Yahoo[4], and product search result page on Amazon[3]). Each participant performed one trial for each combination. The presentation order of *Sasayaki* whispering was counterbalanced and the order of *Sasayaki* jump was fixed because *Sasayaki* jump was observed to be able to significantly reduce the task completion time in our pilot study leading us to focus primarily on the *Sasayaki* whispering conditions. Also, if the jump function was not fixed, then the number of samples for each condition became too small. Recruiting more blind participants for larger samples is also difficult.

The participants were first familiarized with the functions of aiBrowser and *Sasayaki*, including how to adjust the volumes and speech speeds of both voices. This involved a different and special webpage. They were then asked to attempt each task with a time limit or until they had failed three times. We prepared four different webpages with similar test problems for each category of webpage. For all of the tasks, each participant was asked to find specified information on the webpages and to report out loud with a phrase such as "Here it is." The observer would then look at the focus of aiBrowser to determine whether or not it was the correct position. We measured the task completion time from when we asked them to start to when they successfully reported finding the desired information. The keyboard events were recorded by the test system.

Tasks 1 and 2: Reading a news article

These two tasks were performed for article pages on two news websites, **Asahi** and **Nikkei**. The participants were asked to find the first paragraph of the article on the page within three minutes. This task assesses their navigation to the main content of a webpage.

Tasks 3 and 4: Shopping for a product

These two tasks were performed at product pages on two e-commerce Web sites, **Amazon** and **Yahoo**. They were asked to first find the price of the product and then find the button to purchase the product. The total time limit for each pair of tasks was five minutes. This task assessed navigation for the important information on the webpages.

Task 5: Searching for a price in a list of search results

The last task involved product search result pages on an e-commerce Web site, Amazon (referred as **A-search**). This kind of webpage has a list of more than twenty products. The participants were asked to find the product with the highest price among the first ten items in the list. This task tested navigation for a more sophisticated task and was timed up to seven minutes.

Questionnaire and Interview

After the experiment, we used a survey with seven-point Likert items from -3/definitely disagree to +3/definitely agree to compare the test conditions in pairs, NS-NJ against S-NJ and NS-J against S-J. This produced subjective scores for the Web experiences with the *Sasayaki* functions. The following list translates the items from the questionnaire. **NS** means NS-NJ or NS-J and **S** means S-NJ or S-J.

- Compared to **NS**, I found **S** to be easier to use
- Compared to **NS**, I found **S** to be useful
- Compared to **NS**, I had more control using **S**
- Compared to **NS**, I found **S** to be more pleasant
- Compared to **NS**, I felt more sure I would finish with **S**
- Compared to **NS**, I rated **S** output quality as better
- I would use **S** if I had access to it

The page overview functions, the text analysis functions, and the *Sasayaki* whispering voice setting features were all explained before we interviewed the participants about these three *Sasayaki* functions, about ideas for improving the system, and about some other questions.

RESULTS

Task Completion Time

Figure 2 shows the average task completion times with standard errors. The average values were 112 seconds for NS-NJ, 126 seconds for S-NJ, 71 seconds for NS-J, and 65 seconds for S-J. Analysis of variance showed significant primary effects for the *Sasayaki* jump ($F_{1,130} = 65.23$, $p < .001$) and the target website ($F_{4,130} = 51.05$, $p < .001$). There were interaction effects for the jump × the target website ($F_{4,130} = 5.06$, $p < .005$), whispering × the target website ($F_{4,130} = 3.33$, $p < .05$). A post hoc analysis indicated that *Sasayaki* jump function significantly decreased the task completion time and the impact of the

[1] Asahi newspaper: http://www.asahi.com/

[2] Nikkei newspaper: http://www.nikkei.com/

[3] Amazon.co.jp: http://www.amazon.co.jp

[4] Yahoo Japan Shopping: http://shopping.yahoo.co.jp/

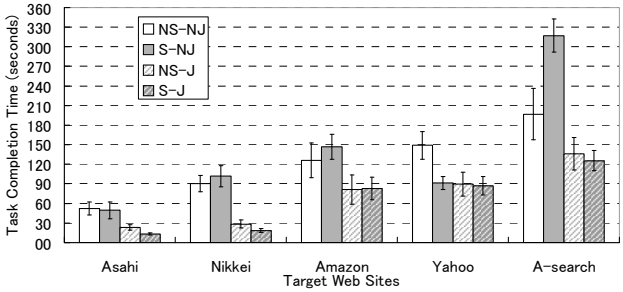

Figure 2. The average task completion times

	Asahi	Nikkei	Amazon	Yahoo	A-search
NS-NJ	0	1	7	2	4
S-NJ	0	0	4	0	3
NS-J	0	0	0	1	0
S-J	0	0	0	0	0

Table 1. Numbers of participants who failed to complete a task within the time limit

	Asahi	Nikkei	Amazon	Yahoo	A-search
NS-NJ	0	0	2	3	1
S-NJ	0	0	0	0	0
NS-J	0	0	0	1	0
S-J	0	0	0	0	0

Table 2. Number of participants had incorrect answers during a test

Sasayaki whispering function depended on the task. The data for the trials that were stopped because of the time limits are not included in this graph. Seven participants ran out of time in the Amazon webpage with NS-NJ and some participants were out of time for Nikkei, Yahoo, and A-search with the NS-NJ condition, for Amazon and A-search with the S-NJ condition and for Yahoo with the NS-J condition. With the S-J condition, there were no time out problems.

Error Rates

Table 1 shows the number of participants who failed to complete tasks due to the time limits. The reason many people failed on the Amazon webpages was that these webpages have a difficult structure. Table 2 shows the number of participants who had incorrect answers during the task. No participant had more than two mistakes and no participant made any mistakes with the S-NJ or S-J conditions.

Subjective Scores

Figures 3 and 4 show the average subjective ratings with standard errors for the seven questions comparing conditions without jump (NS-NJ vs. S-NJ) and conditions with jump (NS-J vs. S-J). All of the participants except P4

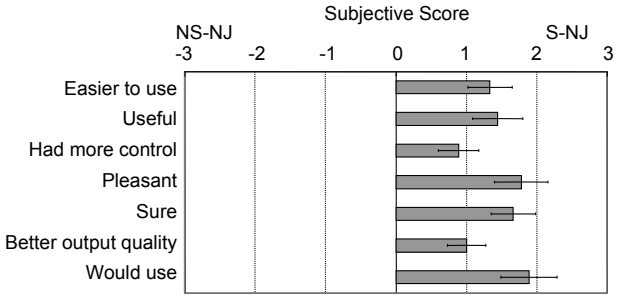

Figure 3. The average subjective scores for NS-NJ vs. S-NJ

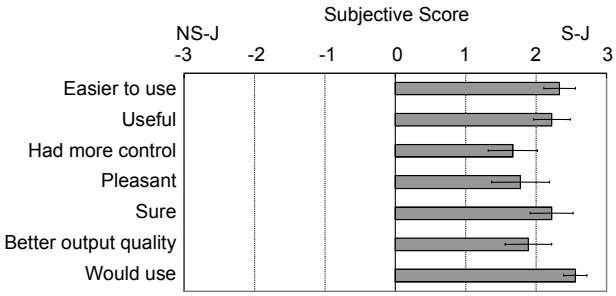

Figure 4. The average subjective scores for NS-J vs. S-J

gave positive ratings. The average score is 1.43 for NS-NJ vs. S-NJ and 2.10 for NS-J vs. S-J. All of the scores of NS-J vs. S-J are equal to or higher than those of NS-NJ vs. S-NJ. "Would use" received the highest score in both, with or without jump comparisons. For conditions without the *Sasayaki* jump feature, the "pleasant" and "sure" questionnaire items received higher scores. For conditions with jump, the "easy to use", "useful", and "sure" items received higher scores.

Navigation Behavior Analysis

The participants finished all of the tasks within 30 to 50 minutes. While working on the tasks they made from 1,800 to 3,500 keystrokes in navigating the webpages. Figure 5 shows the average numbers of keystrokes for each condition, including all of the trials with data. The average values were 194 keystrokes for NS-NJ, 175 for S-NJ, 77 for NS-J, and 69 keystrokes for S-J. Analysis of variance showed significant primary effects for the *Sasayaki* jump ($F_{1,152} = 80.090$, $p < .001$) and for the target website ($F_{4,152} = 38.098$, $p < .001$). There were interaction effects for the jump × the target website ($F_{4,152} = 10.85$, $p < .001$). A post hoc analysis indicated that the *Sasayaki* jump function decreased the numbers of keystrokes. Table 3 shows the average ratios for using the navigation commands. More than 60% of the navigation commands were "Next element" and about 20% of the navigation commands were "Previous element". The rest of navigation commands were mostly for jumps.

Type	NS-NJ	S-NJ	NS-J	S-J
Next element	64.6%	66.7%	61.3%	60.0%
Previous element	21.8%	18.6%	21.1%	19.4%
Next heading	4.0%	6.0%	1.0%	1.5%
Previous heading	1.4%	2.0%	0.5%	0.3%
Next link	3.3%	3.2%	0.3%	1.1%
Previous link	3.1%	2.2%	0.0%	0.4%
Sasayaki jump (next)	0.0%	0.0%	9.4%	10.5%
Sasayaki jump (prev)	0.0%	0.0%	3.9%	5.3%
Others	1.8%	1.3%	2.4%	1.5%

Table 3. The average ratios of navigation command usage

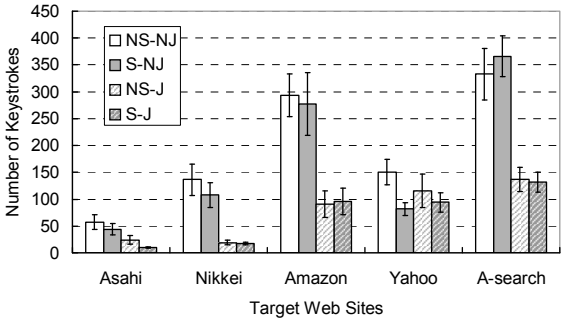

Figure 5. The average number of keystrokes

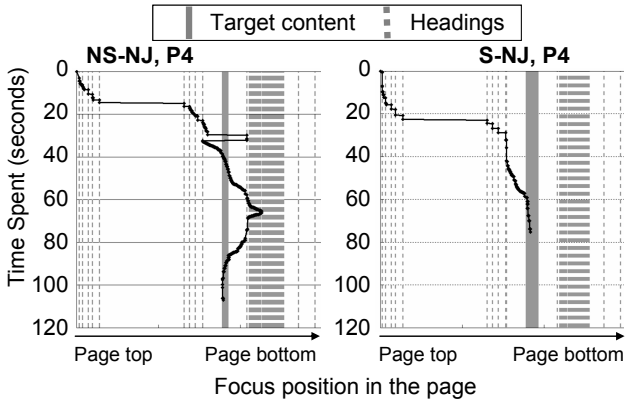

Figure 6. Navigation traces on Nikkei newspaper with NS-NJ and S-NJ by P4.

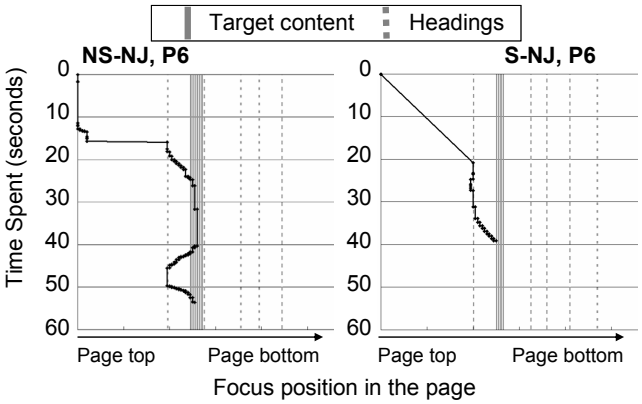

Figure 7. Navigation traces on Asahi newspaper with NS-NJ and S-NJ by P6.

Navigation Trace Analysis

Each blind user has a strategy for Web navigation and may change that strategy based on the type of webpage or the purpose of navigation. Based on their different strategies, the nine participants can be categorized into three groups. In the first group, five participants (P1, P2, P3, P4, P5) mainly used the heading jump and page search as their second choice. In the second group, three participants (P6, P7, P9) mainly used page searching functions and heading jumps as second choice. There was only one participant in group three who mainly used the cursor keys, usually with the tab key, to explore elements linearly or to jump between links (P8).

Figure 6 shows the traces for P4 with the NS-NJ and S-NJ conditions in the Nikkei webpage. The vertical axis is the time spent on each task and the horizontal axis is the focus position in the page. The black line shows how the participant moved on the page and the vertical gray line indicates the area with the target element for that task. The dashed gray lines are for heading elements. In both the traces for NS-NJ and for S-NJ, by using heading jump functions the participants' focus arrived at the same location, which was just ahead of the target area. For NS-NJ, the participant accidentally passed the target area and came back to it. For S-NJ, the *Sasayaki* whispers such as "close to main" and "entering main" would be helpful for the participant in finding the target content. Figure 7 shows another trace of P6 with NS-NJ and S-NJ. This user's navigation strategy was mostly based on page search. For NS-NJ, P6 failed in the search because of a typing mistake. He then used the heading jump function and found the main content. However P6 checked the previous content again to make sure about the goal of the task (in the period from 40 to 50 seconds). This kind of pattern was observed 11 times without *Sasayaki* whispering and only 5 times with *Sasayaki* whispering, indicating that the *Sasayaki* whispering improved users' confidence that they had found the desired webpage elements.

There were also cases in which the *Sasayaki* functions didn't work well. Figure 8 shows the traces for the Asahi newspaper with S-NJ by P3 and P9. P3 heard "close to main" as a *Sasayaki* whisper but P3 could not navigate step by step before passing the target area and then skipped the main content by using heading jump. Unfortunately the *Sasayaki* system basically tries to convey information about the current position rather than about the content the user has already passed. This design was chosen to reduce the amount of information whispered by *Sasayaki* so that the two voices could be heard more easily. P9 was confused by the *Sasayaki* whispering and said "the structure of the page from the *Sasayaki* whispers was not easy to understand" and "'close to main' should mean within two or three items of the position". This case suggests that we need either a

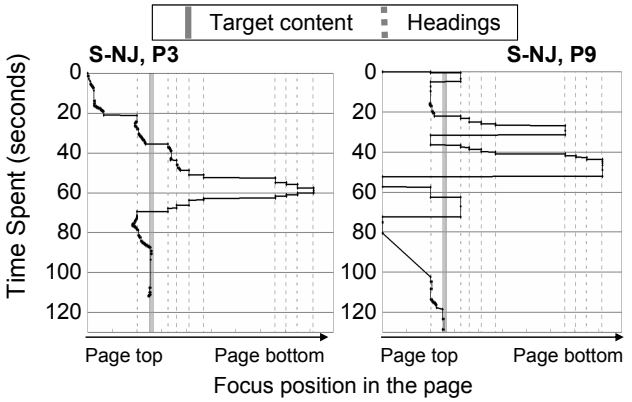

Figure 8. Navigation traces on Asahi newspaper with S-NJ by P3 and P9.

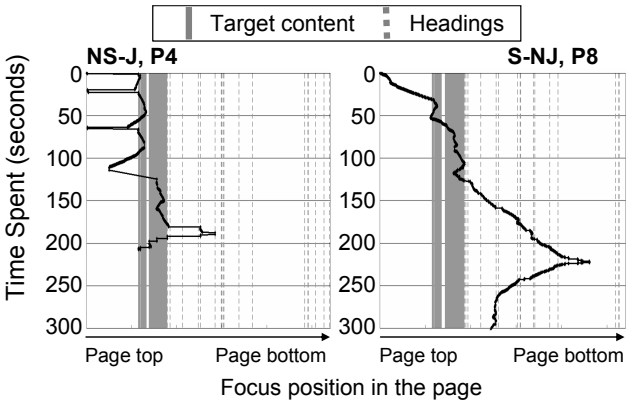

Figure 9. Navigation traces on Amazon product pages with NS-J by P4 and S-NJ by P8.

personalized control for the *Sasayaki* content details, or better training so that users would be more familiar with our *Sasayaki* system.

Another problem involved incorrect assumptions about the structure of the webpage, especially for the Amazon product page. In this page, the "add to cart" button, the product title as the first heading element, and the price appeared in that order. However, most participants thought that the "add to cart" button must be below the price. Figure 9 shows the traces of Amazon webpages for NS-J by P4 and for S-NJ by P9. There are two target areas in this task. The first area contains the "add to cart" button and the other contains the price information. P4 found the "add to cart" button by using the *Sasayaki* jump function, and then tried to search for the price in front of the button three times. Finally P4 used the page search function to get the price. In contrast, P8's strategy is to explore the webpage elements one by one, so P8 was aware of the position of the "add to cart" button when the *Sasayaki* whisper said "often used button" to emphasize it. P8, however, searched for the "add to cart" button after the price and thus could not finish.

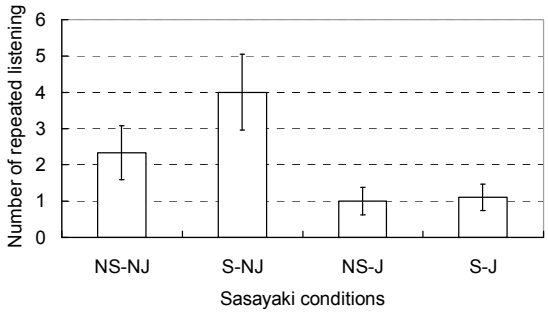

Figure 10. The average number of repetitions on Amazon search result pages.

DISCUSSION

Navigation Improvements with Sasayaki

Even though most of the participants gave positive ratings for the test conditions with *Sasayaki* whispers compared to those without *Sasayaki*, the quantitative results showed no significant difference between the conditions with and without *Sasayaki* whispers regarding the task completion time or number of keystrokes. The navigation traces also suggest improvements of the navigation behaviors. One of interpretations of the results is that assisted voice (*Sasayaki*) gave participants "feeling of confidence", but the processes of mental model building was not enhanced enough to improve the performance in the tasks.

The participants' comments support this interpretation. Typical positive comments include "I feel confident with *Sasayaki* compared to without *Sasayaki*" (P7), "A *Sasayaki* whisper, e.g. 'close to main' is so nice, making me feel comfortable" (P6) and "It is useful for exploring the structure of unfamiliar webpages" (P8). Those comments suggested that the *Sasayaki* system reduced the stresses of unaided navigation and the strain of sustained listening.

We also see evidence in the traces of navigation behaviors. As shown in Figure 6, the browsing pattern of passing the target and coming back to increase their confidence were frequently recognized in the conditions without *Sasayaki* whispering. Blind users concentrated on the synthesized voices to track their positions in the page. This indicates that unaided navigation is a very hard task that lowers the user's confidence. A *Sasayaki* system can help to increase their confidence with relatively little concentration on the second voice. The fact that the number of incorrect answers by the participants was reduced by *Sasayaki* whispering is evidence that supports this finding.

Based on the extracted approaches from the less successful patterns, we also got hints on how to improve the design of *Sasayaki* voices to enhance the mental model building processes. The cases shown in Figures 8 and 9 could be addressed by presenting a structure map for the users to study before they start to navigate on the page. In fact, some participants requested such a function in the interview

session and one participant suggested that providing a tree view would be useful for him to form a basic idea of the webpage structure before exploring.

Are Two Simultaneous Voices Distinguishable?

Although we didn't explicitly experiment with the effects of the two simultaneous voices, we recognized repeated listening as a noticeable behavior of the participants in the Amazon search result pages. Figure 10 shows the average numbers of repetitions within each *Sasayaki* condition. For the webpages where they were asked to report the index of the item with the highest price among the top ten items, the *Sasayaki* whisper was available for the first element of each result item and the first element was the index number. In the S-NJ condition, seven participants did listen to the rank repeatedly. In contrast, we observed a smaller number of repeated listening in the S-J condition, as shown in Figure 10. In the test conditions with *Sasayaki* jump features (NS-J and S-J), the participants seem to be able to recognize the two voices simultaneously and properly interpret the *Sasayaki* whispering. This result would be caused by some learning effects as in the experiment all participants tested the conditions without *Sasayaki* jump features first.

The participants did not complain about audibility of the secondary voice and felt confident about it. Two participants commented about the two parallel voices, "the two physically separated voices were sufficiently distinguishable" (P2, P8), and "it could be improved by adjusting the volumes and combinations of the types of voices" (P6). This result might be due to the fact that blind people have good listening skills to compensate for their loss of visual perception. However even sighted people can recognize multiple voices near them. Further studies on improving the listenability of the secondary voice might be necessary, for example by adjusting the voice output timing and choosing more distinguishable combinations of the two voices. The information density of the auditory user interface could be increased by the *Sasayaki* approach. This would be a paradigm shift for the auditory interface.

Potential Applications

As we noted, *Sasayaki* is a general concept for a user interface that provides supplemental information for various kinds of people via an audio channel. This concept could be applied to many real world environments. One example might be navigation for blind pedestrians. This kind of system could be enhanced with a secondary voice to augment the primary voice that provides walking directions, for example to provide user-generated information about nearby restaurants. The creation of supplemental audio information to provide situational support has been already introduced in some emerging technologies (e.g., [36]). Telephony applications could also be enhanced to increase information density. For example, when a user is trying to access a telephony application for the first time, the user could hear the main voice, and, at the same time, if the system detects the user status to be "needs help" by analyzing the user's behavior and contextual information, then a second voice could provide situational instructions or tips on how to navigate or interact with the application more easily.

CONCLUSION

This paper describes a concept called *Sasayaki* which augments a primary voice output with a secondary voice that whispers contextually relevant information automatically or in response to user requests. A prototype system was implemented as a plug-in system for a voice-based Web browser with a small API for the *Sasayaki* controls. An empirical evaluation with nine visually impaired users showed that the *Sasayaki* system significantly improved their navigation behaviors and increased their confidence levels. The jump function based on *Sasayaki* significantly increased the navigation performance. The results also show the possibilities of the *Sasayaki* approach, with two simultaneous voices increasing the information density of the auditory user interface. For future work we will do more studies to explore the advantages of simultaneous voices, study the emotional effects of the *Sasayaki* functions, and also test *Sasayaki* for applications for other population groups.

REFERENCES

1. Internet World Stats, World Internet Users and Population Stats, http://www.internetworldstats.com/stats.htm.
2. UNESCO, International Literacy Statistics: A Review of Concepts, Methodology and Current Data, http://www.uis.unesco.org/template/pdf/Literacy/LiteracyReport2008.pdf.
3. WHO, Fact sheet of visual impairment and blindness, http://www.who.int/mediacentre/factsheets/fs282/en/.
4. ITU. Measuring the Information Society 2010. http://www.itu.int/ITU-D/ict/publications/idi/2010/.
5. Takagi, H., Saito, S., Fukuda, K. and Asakawa, C. Analysis of navigability of Web applications for improving blind usability. *ACM Trans. Comp.-Hum. Interact* 14:3 (2007), 13.
6. Barnicle, K. Usability testing with screen reading technology in a Windows environment. In *Proc. CUU 2000*, ACM Press (2000), 102-109.
7. Lazar, J., Allen, A.and Kleinman, J. and Malarkey, C. What Frustrates Screen Reader Users on the Web: A Study of 100 Blind Users. *International Journal of Human-Computer Interaction* 22:3 (2007), 247-269.
8. Maes, P. 1994. Agents that reduce work and information overload. *Communications of. ACM* 37, 7 (1994), 30-40.
9. Bederson, B. B. Audio augmented reality: a prototype automated tour guide.,In *Proc. CHI 1995*, ACM Press, (1995), 210-211.

10. Sawhney, N. and Schmandt, C. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Trans. Comput.-Hum. Interact.* (7:3), (2000), 353-383.

11. Eckel, G. Immersive Audio-Augmented Environments: The LISTEN Project, *Fifth Framework Programme, Creating a user-friendly information society* (IST), (2001), 571.

12. Kalantari, L., Hatala, M. and Willms, J. Using semantic web approach in augmented audio reality system for museum visitors. In *Proc. WWW 2004*, ACM Press, (2004), 386-387.

13. Miyashita, T., Meier, P., Tachikawa, T., Orlic, S., Eble, T., Scholz, V., Gapel, A., Gerl, O., Arnaudov, S. and Lieberknecht, S. An Augmented Reality museum guide. In *Proc. ISMAR 2008*, IEEE Computer Society (2008),103-106.

14. Shoval, S., Borenstein, J. and Koren, Y. The Navbelt - A Computerized Travel Aid for the Blind Based on Mobile Robotics Technology. *IEEE Trans. on Biomedical Engineering* (45:11) (1998), 1376-1386.

15. Jones, M., Jones, S., Bradley, G., Warren, N., Bainbridge, D. and Holmes, G. ONTRACK: Dynamically adapting music playback to support navigation. *Personal and Ubiquitous Computing* (12:7), (2008), 513-525.

16. Stylos, J., Myers, B. A. and Faulring, A. Citrine: providing intelligent copy-and-paste. In *Proc. UIST 2004,* ACM Press (2004), 185-188.

17. Wagner, E. J. and Lieberman, H. Supporting user hypotheses in problem diagnosis. In *Proc. IUI 2004,* ACM Press (2004), 30-37.

18. Roth, P., Petrucci, L., Pun, T., and Assimacopouls, A. Auditory browser for blind and visually impaired users, In *Proc. CHI 1999*. ACM Press (1999), 218-219.

19. Yu, W., McAllister, G., Strain, P., Kuber, R. and Murphy, E. Improving web accessibility using content-aware plug-ins. In *Proc. CHI 2005,* ACM Press (2005), 1893-1896.

20. Dontcheva, M., Drucker, S. M., Wade, G., Salesin, D., and Cohen, M. F. Summarizing personal web browsing sessions. In *Proc. UIST 2006*. ACM Press (2006), 115-124.

21. Hartmann, M., Schreiber, D. and Mühlhдuser, M. AUGUR: providing context-aware interaction support. In Proc. of symposium on Engineering interactive computing systems, ACM Press (2009), 123-132.

22. Parente, P.. Clique: a conversant, task-based audio display for GUI applications. SIGACCESS Access. Comput. 84 (2006), 34-37.

23. Mahmud, J. U., Borodin, Y., and Ramakrishnan, I. V. Csurf: a context-driven non-visual web-browser. In *Proc. WWW 2007.* ACM Press (2007), 31-40.

24. Borodin, Y., Bigham, J. P., Raman, R. and Ramakrishnan, I. V. What's new? Making web page updates accessible. In *Proc. ASSETS 2008,* ACM Press (2008), 145-152.

25. Yesilada, Y., Stevens, R., Harper, S. and Goble, C. Evaluating DANTE: Semantic transcoding for visually disabled users. *ACM Trans. Comput.-Hum. Interact.* (14:3), (2007), 14.

26. Harper, S. and Patel, N. Gist summaries for visually impaired surfers. In *Proc. ASSETS 2005,* ACM Press (2005), 90-97.

27. Miyashita, H., Sato, D., Takagi, H. and Asakawa, C. aiBrowser for multimedia: introducing multimedia content accessibility for visually impaired users. In *Proc.ASSETS 2007,* ACM Press (2007), 91-98.

28. Lunn, D., Bechhofer, S. and Harper, S. The SADIe transcoding platform. In Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A), ACM Press (2008), 128-129.

29. Takagi, H., Kawanaka, S., Kobayashi, M., Itoh, T., and Asakawa, C. Social accessibility: achieving accessibility through collaborative metadata authoring. In *Proc. ASSETS 2008.* ACM Press (2008), 193-200.

30. Chen, C. L. and Raman, T. V. AxsJAX: a talking translation bot using google IM: bringing web-2.0 applications to life. In *Proc. the 2008 international cross-disciplinary conference on Web accessibility (W4A)*, ACM Press (2008), 54-56.

31. Goble, C., Harper, S., and Stevens, R. The travails of visually impaired web travellers. In *Proceedings of the Eleventh ACM on Hypertext and Hypermedia.* ACM Press (2000), 1-10.

32. Eclipse ACTF Accessibility Internet Browser, http://www.eclipse.org/actf/downloads/tools/aiBrowser/.

33. Kawanaka, S., Borodin, Y., Bigham, J. P., Lunn, D., Takagi, H. and Asakawa, C. Accessibility commons: a metadata infrastructure for web accessibility. In *Proc. ASSETS 2008,* ACM Press (2008), 153-160.

34. Kanayama, H., Nasukawa, T. and Watanabe, H. Deeper sentiment analysis using machine translation technology. In *Proc. of the 20th international conference on Computational Linguistic',* Association for Computational Linguistics (2004), 494.

35. Shaojian Zhu, Daisuke Sato, Hironobu Takagi, and Chieko Asakawa. Sasayaki: an augmented voice-based web browsing experience. In *Proc. ASSETS 2010.* ACM Press (2008), 279-280.

36. Wilson, J., Walker, B. N., Lindsay, J., Cambias, C., and Dellaert, F. SWAN: System for Wearable Audio Navigation. In *Proc. ISWC 2007.* IEEE Computer Society (2007), 1-8.