# Sasayaki: An Augmented Voice-based Web Browsing Experience


Shaojian Zhu
UMBC
1000 Hilltop Circle
Baltimore, Maryland, USA
+1 410 455 3883
szhu1@umbc.edu

Daisuke Sato
IBM Research – Tokyo
1623-14 Shimo-tsuruma
Yamato, Kanagawa, Japan
+81 46 215 4793
dsato@jp.ibm.com

Hironobu Takagi
IBM Research – Tokyo
1623-14 Shimo-tsuruma
Yamato, Kanagawa, Japan
+81 46 215 4557
takagih@jp.ibm.com

Chieko Asakawa
IBM Research – Tokyo
1623-14 Shimo-tsuruma
Yamato, Kanagawa, Japan
+81 46 215 4633
chie@jp.ibm.com



## ABSTRACT

While the usability of voice-based Web navigation has been steadily improving, it is still not as easy for users with visual impairments as it is for sighted users. One reason is that sequential voice representation can only convey a limited amount of information at a time. Another challenge comes from the fact that current voice browsers omit various visual cues such as text styles and page structures, and lack meaningful feedback about the current focus. To address these issues, we created Sasayaki, an intelligent voice-based user agent that augments the primary voice output of a voice browser with a secondary voice that whispers contextually relevant information as appropriate or in response to user requests. A prototype has been implemented as a plug-in for a voice browser. The results from a pilot study show that our Sasayaki agent is able to improve users' information search task time and their overall confidence level. We believe that our intelligent voice-based agent has great potential to enrich the Web browsing experiences of users with visual impairments.



## Categories and Subject Descriptors

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

Design, Human Factors

## Keywords

Sasayaki, Voice Agent, Voice Browser, Web Accessibility.


## 1. INTRODUCTION

The Web has brought great changes to many people's lives, even including people with visual impairments. Voice-based user interfaces have expanded their possibilities and capabilities in navigating the Web. Therefore enormous efforts have been dedicated to improving voice-based Web navigation. However, voice-based navigation is still difficult for the users, especially in regard to grasping overviews of webpages and tracking their own positions within the webpages.




*ASSETS'10*, October 25-27, 2010, Orlando, Florida, USA.

ACM 978-1-60558-881-0/10/10.


Takagi et al. reported that voice browser users found it difficult and very time-consuming to navigate to certain elements of webpages [5]. They also found that because screen reader users are worried about missing any important information on a page, they usually prefer to scan each element, which essentially wastes a lot of time on the elements of less important sections. Lazar et al. reported that screen reader users become frustrated for many reasons, including the technical difficulties of navigating, and this negatively affects their feelings and their performance [4].

Two reasons for these difficulties in navigation have been identified. First, voice browsers can deliver only limited chunks of information at a time and in a transient sequential way, and the transformations from the original webpages therefore ignore or mask various visual cues such as the page layout information. This loss of information makes it almost impossible for voice browser users to easily form a meaningful concept of the overall webpage structure, especially for complicated webpages. The other reason is that voice browsers do not provide sufficient contextual support. Unlike sighted people who can always find their cursor position and the local context information at a glance, the voice browser users must always remember where they are and where they have been. Currently, the supportive contextual information is fundamentally inaccessible to the ear.

In order to improve their voice browsing experience, we have created Sasayaki (meaning "whisper" in Japanese), an intelligent voice agent, which helps voice browser users overcome these difficulties and fill in the gaps in their navigation experiences compared to the experiences of sighted people.

## 2. SASAYAKI AGENT

The Sasayaki agent is designed to "whisper" necessary information through a second voice channel. This enables the users to retain awareness of their current context and obtain overviews of webpages. Sasayaki tries to act in a supportive role to help the users like a sighted person to assist the visually impaired user (Figure 1). It tracks both the status of the voice browser and the behaviors of the user. Then Sasayaki can determine the most appropriate contextual or task-relevant support. This supportive information then becomes available to the users through the whispering channel, either automatically or as requested. We have implemented a prototype of the Sasayaki agent system. It is designed as a plug-in component for a voice browser called aiBrowser [1].

In order to provide contextual support, Sasayaki retrieves pre-defined role-based data about webpages from our Accessibility Commons [3] server. The typical roles for content include 'main',

'header', 'advertisement', and so on. The Sasayaki plug-in also monitors the position of the virtual cursor in the voice browser and uses key events to track user behaviors. The plug-in can then create the most appropriate information that will help the user. For example, when the user reaches the defined main content body for the first time, the system uses "Entering main" as a contextual prompt. The agent also allows users to deliberately jump between these role-based elements.

In addition, for webpages in the same category, important information is often shared. For example, each of the product pages in an online shopping website has certain important data, such as price, in-stock status, and shipping information. It often takes time and effort for a user to find that information on a complex webpage, but Sasayaki can collects that data based on the page's category and output it in a convenient format. This makes it easy for the user to use such categorized webpages.

Another problem for voice browsers users involves understanding the reviews in online shopping webpages. In this situation, Sasayaki can invoke a text-mining system [2] to analyze and summarize the sentiments in a large number of reviews. Sasayaki retrieves the original user reviews from the webpages and passes them to the text-mining system. Sasayaki then whispers statistical summaries of the sentiments in the reviews. This helps the user save time and avoids the work of listening to a large amount of text and then creating a mental summary.

As Sasayaki finds information that may be helpful to the user, it uses a secondary voice to announce it. This comes from a different sound device and uses a different speech synthesizer engine, thus creating the whispered-hints scenario. In this way, Sasayaki efficiently and simultaneously presents both voices to the user, making Sasayaki less intrusive but still effective.

## 3. PILOT STUDY

Three visually impaired users (two blind and one low vision, aged 37 to 44) took part in our pilot study of the Sasayaki prototype. All of them are native speakers of Japanese with prior JAWS experience, but none of them had used aiBrowser. We used two versions of aiBrowser in these experiments, the original aiBrowser and a Sasayaki-enhanced aiBrowser.

During the study, the participants were asked to perform the same types of tasks with each version of aiBrowser. They were directed to a product page from a well known Japanese shopping website, and were asked to find specific information. They were also directed to a news webpage on a prominent Japanese news portal website and were asked to read the main content before finding a particular link. At the end of the study, the participants responded to a questionnaire with demographic information, satisfaction ratings, and their comments about the Sasayaki agent.

The Sasayaki agent system was able to effectively improve the users' navigation. Users with Sasayaki spent much less time navigating to the targeted page elements. By using some of the advanced Sasayaki functions such as page overview and role-based jumping, the voice browser users were able to retrieve the required information quite rapidly. Sasayaki also made it easier for the users to complete the tasks.

All of the participants gave consistently high ratings for the usefulness and comfort of Sasayaki's role-based context support.

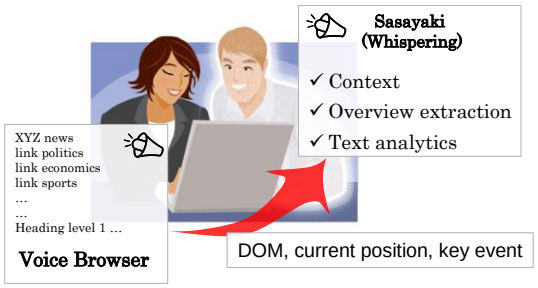


Figure 1. Concept of Sasayaki User Interface Agent

The expert participant was neutral towards our advanced content analysis functions, which may indicate that expert users already have higher confidence in their own expertise and experience in finding the page elements. However, both of the less experienced participants found the advanced functions to be very helpful and rated them highly. All three participants were eager to use Sasayaki in future. They also showed high confidence in their abilities to do Web browsing with Sasayaki. One user did report that the use of two simultaneous voices was confusing, but the others did not find this problematic. Based on these reports, we think we should add user preference settings to control the Sasayaki timing to improve the user experience.

## 4. CONCLUSION

We are prototyping Sasayaki, an intelligent voice-based user agent to improve the quality of voice browser navigation. A pilot study showed that Sasayaki can greatly improve users' navigation experiences. A more formal study will be conducted to more carefully evaluate the potential usefulness of our Sasayaki system.

## 5. REFERENCES


[1] ACTF Accessibility Internet Browser for Multimedia (aiBrowser) http://www.eclipse.org/actf/downloads/tools aiBrowser/index.php, retrieved 07-01-2010

[2] Kanayama, H. and Nasukawa, T. 2008. Textual demand analysis: detection of users' wants and needs from opinions. In *Proc. of the 22nd int. Conf. on Computational Linguistics* – Vol. 1. 409-416.

[3] Kawanaka, S., Borodin, Y., Bigham, J. P., Lunn, D., Takagi, H., and Asakawa, C. 2008. Accessibility commons: a metadata infrastructure for web accessibility. In *Proc. of the 10th int. ACM SIGACCESS Conf. on Computers and Accessibility.* Assets '08. 153-160.

[4] Lazar, J., Feng, J., and Allen, A. 2006. Determining the impact of computer frustration on the mood of blind users browsing the web. In *Proc. of the 8th int. ACM SIGACCESS Conf. on Computers and Accessibility*. Assets '06. 149-156.

[5] Takagi, H., Saito, S., Fukuda, K., and Asakawa, C. 2007. Analysis of navigability of Web applications for improving blind usability. *ACM Trans. Comput.-Hum. Interact.* 14, 3 (Sep. 2007), 13.