SoundGlance: Briefing the Glanceable Cues of Web Pages for Screen Reader Users

Jaeyoon Song*
Seoul National University
Seoul, South Korea
song@jaeyoon.io

Kiroong Choe*
Seoul National University
Seoul, South Korea
krchoe@hcil.snu.ac.kr

Jaemin Jo
Seoul National University
Seoul, South Korea
jmjo@hcil.snu.ac.kr

Jinwook Seo
Seoul National University
Seoul, South Korea
jseo@snu.ac.kr

*Both authors contributed equally to the paper

ABSTRACT

Screen readers have become a core assistive technology for blind web users to browse web pages. Although screen readers can convey the textual information or structural properties of web pages, they cannot deliver their overall impression. Such a limitation hinders blind web users from obtaining an overview of the website, which non-blind people can do in a short time. As such, we present SoundGlance, a novel application that briefly delivers an auditory summary of web pages. SoundGlance supports the screen reader users by converting the important glanceable cues of the pages into sound. The feasibility of prototype was examined in a pilot study with fourteen blind people. Several practical insights were derived from the experiment.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3312865

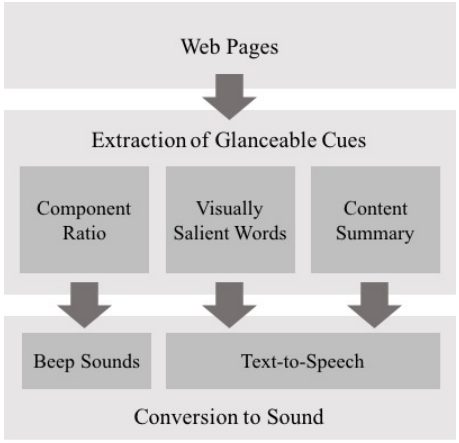

Figure 1: How SoundGlance works

KEYWORDS

Accessibility; Screen readers; Visually-Impaired

INTRODUCTION

Modern web pages are often complex and thus it takes time to fully comprehend them. A single glance on such a website provides a number of useful cues including the dominance of advertisements, visually salient words, and skimmable content, of which many web users tend to make good use.

However, such "glanceable" cues have poor accessibility for the visually impaired users. Most of the visually impaired people use screen readers to browse web pages, but screen readers only allow sequential access to textual information of a website. Although there are complementary features such as site summary, they rely heavily on the HTML structure of the website and simply reads aloud the headings, which a lot of times does not provide proper information. As a result, a lot of screen reader users end up with inefficient and time-consuming experience [5].

To address these challenges, we introduce SoundGlance, a system for screen reader users that turns the glanceable cues of a website into sound. Specifically, SoundGlance collects the following information: (1) dominant components (e.g., image, text, ad); (2) visually salient words or phrases; and (3) an overview of the content. Then the information is converted to beep sounds and speech as short as a glance. We conducted a pilot study (n=14) to test the feasibility of our system, from which design considerations for the future redesign were derived.

BACKGROUND

Browsing Strategies with Screen Readers

Screen reader users tend to employ browsing strategies such as navigating using heading tags and shortcuts to the main content if provided [2]. Along with reviewing literature, we investigated current strategies by communicating with real-world users. We conducted a preliminary observational study with a visually impaired person who is highly skilled in using screen readers. The participant was asked to browse a given website and notify us when a decision was made whether the website provides good quality information about a given topic. During the study, the participant started listening to the title, headings, and then sequentially listened to the main content. It took from 30 seconds to several minutes for the participant to make the decision. The web pages that took less than a minute mostly provided a shortcut to the main content or contained a heading directly related to the topic. However, for other pages that did not provide such conveniences, which is in fact the case for most websites, it took several minutes.

Currently, screen readers have few features to address these delays. For instance, the JAWS screen reader provides a "skim reading" function that allows users to read the first line or sentence of each

paragraph, and a site summary that informs about the number of headings, links, tables, and image tags, followed by reading the title, navigation, and headings [7]. Nevertheless, these features have limitations in that they merely rely on the markup provided by the developer, resulting in confusion when the markup poorly follows the accessibility guidelines.

Glanceable Cues

Glanceable cues are visual and non-visual signals about the website, which could be gained in a short moment and facilitate more efficient web surfing. According to Wathen and Burkell's model for how users assess the online information [8], evaluation of surface credibility and message credibility takes place prior to the full review of the content. Based on the model, we investigated what kind of glanceable cues are there and which ones should be included in the prototype.

First, surface evaluation is an assessment of surface characteristics such as appearance (e.g., color tone, graphics), usability (e.g., interactivity), and organization of information (i.e., layout). These surface characteristics are known to play an important role in credibility evaluation of websites [4]. Among surface characteristics, however, the aesthetic cues such as color tone could cause confusion since they are too subjective to be delivered accurately. Rather, the information about whether the website is image-dominant, text-dominant, or full of advertisements along with the visually salient words is comparatively objective measurement and holds a significance in terms of web accessibility.

Second, message evaluation is an evaluation of the overall relevance of the content without thoroughly going through the whole web page. Ahmed et al. [1] suggested making the process of message evaluation accessible for the screen reader users by designing a "viable algorithm for automatic summarization". Similarly, SoundGlance also provides the content summary using natural language processing in order to facilitate the user's decision about the message of the website.

PROTOTYPE DESIGN OF SOUNDGLANCE

Web Page Selection

We implemented SoundGlance as an offline application and tested it with selected web pages. Based on the trend report of Google and Naver–a major Korean search engine–six popular keywords were chosen as shown in Table 3. For each keyword, 30 URLs were collected from the Google search results of the keyword, i.e., top and bottom fifteen among top 100 results [6]. Assuming that the goal is to obtain good quality information related to the keywords, three authors of this paper rated all the 180 web pages based on how much it is worth to achieve the given goal, using a five-point scale. Before doing this, the authors made sure to reach a consensus on the interpretation of the criteria, referring to the literature [6]. To enhance the reliability, the three rating sets were normalized and averaged, and eight pages per keyword that had the most consistent ratings were finally selected.

Table 1: The following are six popular keywords used for the web page selection. The keywords were selected based on the trend report of major search engines. Then, the web pages were chosen among the top 100 search results of each keyword at the most commonly used search engine.

	Keyword
1	Typhoon preparation tips
2	Attractions in Bangkok
3	How AlphaGo works
4	Konjac benefits
5	MSG side effects
6	Psy comeback

Figure 2: Interface we used to generate the dataset for CNN training

Extraction of Glanceable Cues

Ratio of Text, Image, and Advertisements. Whether a website is mainly composed of images (e.g., Instagram), texts (e.g., Wikipedia), or advertisements does not simply correspond to the number of HTML tags. Hence, a pre-trained convolutional neural network (CNN) was used to automatically calculate the dominance of each component of a website. To collect vision-based data, we collected screenshots of 39 web pages and manually marked the area of each component. Each marked area was separated as a patch, annotated with the computed style properties. The network was trained on the data of the generated patches. The validation had an accuracy of 87.38%.

Visually Salient Words. Important words in a web page are highlighted to be visually salient. However, screen readers often fail to detect important words owing to various underlying implementations resulting in the same look. For example, bold text can be implemented with not only tags, but also <b> tags or stylesheets, although the latter is often not detected. SoundGlance collects all text segments in a page and finds the five most salient words with regard to font weights and sizes.

Content Summary. Although skimming is a commonly used strategy to grasp the content, screen readers sequentially read the content, leading to its inefficiency. To minimize such delays, SoundGlance automatically synthesizes a summary from the text of a web page. The extraction is performed using the LexRank algorithm, a natural language processing algorithm that computes the most central sentences in a document [3]. SoundGlance selects five sentences presumed to be most important, thus, providing an overview of the main content.

Conversion to Sound

The ratio of page components was sonified into a series of beep sounds, where the type and area of each component was mapped to the pitch and length, respectively [9]. Specifically, ads were represented with high-pitch sounds to provide a negative impression, images and texts were represented with middle and lowest pitched sounds, respectively. Visually salient words and content summary were delivered as speech by using a text-to-speech conversion.

PILOT STUDY

Design and Methods

We had fourteen visually impaired people participate in the pilot study of SoundGlance. All the participants had a previous experience in using screen readers to browse web pages. The study was conducted in a computer lab and consisted of six sessions and an online questionnaire. Each session lasted approximately five minutes and used eight stimuli (4 keywords × 2 stimuli each). For the stimuli, along with the audio files generated by SoundGlance for 48 selected web pages (SG), we also

Table 2: Composition of site summary and SoundGlance

Site Summary
The number of headings (<h1>,...,<h6>) The number of <a> tags The number of <table> tags The number of <img> tags
The content of <title> tag The content of <nav> tag The content of headings (<h1>,...,<h6>)

SoundGlance
Ratio of text, image, and advertisements in the form of beep sounds
Visually salient words
Content summary

Table 3: Average of user responses

Ground Truth	User Response
High	3.30 / 5.00
Middle	2.96 / 5.00
Low	2.79 / 5.00
Total	3.02 / 5.00

constructed the site summary based on what was provided by the Korean JAWS screen reader for each of the same 48 web pages (SS). Half of the six sessions used SG, and another half utilized SS as the stimuli. In total, each participant was exposed to either SG or SS of 48 heterogeneous web pages. We randomized the order of stimuli within a session and the order of the sessions themselves.

The participants were informed about the keyword used before being exposed to each stimulus, assuming that their goal is to obtain good quality information that is related to the given keyword. After listening to the stimulus, the participants were asked to score how much they think it is worth to more thoroughly explore the given web page. When all sessions were completed, they answered the survey regarding the usability of SoundGlance.

Results

The scores from the results were compared to the "ground truth" ones scored by the three authors. The results of SoundGlance showed a weak but positive correlation (r = 0.14). For the usability survey, only 18.8% responded negatively to the following statement: "I will frequently use SoundGlance in the future." These results indicate the potential usefulness of SoundGlance. The following design considerations were derived from the analysis.

C1. Being independent from the markup was a major strength of SoundGlance. Compared with the results of site summary, SoundGlance performed better with the web pages that have poor accessibility support from the developer. For example, several pages used heading tags without consistency or used tags for titles, causing the site summary to convey unimportant information. However, since SoundGlance is independent from the ready-made HTML markup and also considers the non-heading texts that are visually salient, it was able to obtain important information despite its poor markup.

C2. SoundGlance showed high efficiency. The positive comments on SoundGlance were mainly regarding the efficiency. Five out of fourteen responses mentioned SoundGlance helped them quickly obtain the information, emphasizing that it was able to "immediately grasp the overview." On the other hand, positive comments about the site summary were mostly about the familiarity of types of the information and speech sounds, none of which suggested the aspect of efficiency.

C3. The participants were not used to information beyond the HTML structure. Although more people were positive than negative about the overall usability of SoundGlance, its ease of use and learnability obtained the lowest score. Specifically, the users who preferred site summary pointed out familiarity as the main reason–one of them answered that site summary was better in that "it reads the components such as headings in a usual way." This comment suggests that SoundGlance should highly consider balancing the existing strategies with new approaches from the perspective of screen reader users.

CONCLUSION AND FUTURE WORK

This study suggests the screen readers to provide richer information beyond the ready-made HTML markup, and seeks for ways to support screen reader users to browse web more efficiently. SoundGlance is a novel tool that aims to provide such support by conveying the glanceable cues of the website.

The pilot study helped us identify areas for improvement. While C1 and C2 imply the potential benefits of conveying the glanceable cues, C3 presents a challenge. Also, the correlation found between the baseline and user scores needs to be improved, which indicates that SoundGlance might be suitable for only some contexts and needs training prior to being tested. For future work, we will iterate the design of SoundGlance based on the challenges derived from the pilot study. Several participants pointed out it was hard to differentiate each sound, which suggests the necessity to adopt a different sonification approach that takes into account the trade-off between the clear delivery and pleasantness of the sound. We also plan to add a detailed tutorial in order to enhance the learnability. Additionally, we will implement SoundGlance as a browser extension or a screen reader add-on for real-time use.

REFERENCES

[1] Faisal Ahmed, Yevgen Borodin, Andrii Soviak, Muhammad Islam, I.V. Ramakrishnan, and Terri Hedgpeth. 2012. Accessible Skimming: Faster Screen Reading of Web Pages. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 367–378. https://doi.org/10.1145/2380116.2380164

[2] Yevgen Borodin, Jeffrey P. Bigham, Glenn Dausch, and I. V. Ramakrishnan. 2010. More Than Meets the Eye: A Survey of Screen-reader Browsing Strategies. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A) (W4A '10)*. ACM, New York, NY, USA, Article 13, 10 pages. https://doi.org/10.1145/1805986.1806005

[3] Günes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research* 22 (2004), 457–479.

[4] B. J. Fogg, Jonathan Marshall, Othman Laraki, Alex Osipovich, Chris Varma, Nicholas Fang, Jyoti Paul, Akshay Rangnekar, John Shon, Preeti Swani, and Marissa Treinen. 2001. What Makes Web Sites Credible?: A Report on a Large Quantitative Study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*. ACM, New York, NY, USA, 61–68. https://doi.org/10.1145/365024.365037

[5] João Guerreiro. 2016. Towards Screen Readers with Concurrent Speech: Where to Go Next? *SIGACCESS Access. Comput.* 115 (June 2016), 12–19. https://doi.org/10.1145/2961108.2961110

[6] Julia Schwarz and Meredith Morris. 2011. Augmenting Web Pages and Search Results to Support Credibility Assessment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1245–1254. https://doi.org/10.1145/1978942.1979127

[7] Freedom Scientific. [n. d.]. JAWS screen reader. http://www.freedomscientific.com/

[8] C. Nadine Wathen and Jacquelyn Burkell. 2002. Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology* 53, 2 (2002), 134–144. https://doi.org/10.1002/asi.10016

[9] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. 2008. Data Sonification for Users with Visual Impairment: A Case Study with Georeferenced Data. *ACM Trans. Comput.-Hum. Interact.* 15, 1, Article 4 (May 2008), 28 pages. https://doi.org/10.1145/1352782.1352786