Design of Web Pages Accessible by Voice Browsers

Yoko Asano†, *Masahiro Watanabe, and Aya Okano*

Abstract

Since the Web is primarily a visual medium, visually impaired users currently face particularly severe access barriers. This paper introduces a Web design case study of using voice browsers to access Web content.

1. Introduction

As the Internet expands, a broader range of users are using the Web for an increasing variety of applications in a growing diversity of environments. In particular, the Web is becoming an essential point of social contact for users with few opportunities to venture out, such as the elderly and disabled. However, the elderly and disabled often have greater difficulty using the Web than ordinary users due to a lack of coordination, knowledge, skills, and the like. In particular, visually impaired users face severe difficulties obtaining information from Web content that involves the use of graphical user interface (GUI) elements and the representation of data in visual form. Among the visually impaired, most blind and weak-sighted users gather information by using voice browsers to read out the content of Web pages. However, information is lost from non-textual content that cannot be read out easily, making the content difficult to understand. Accordingly, it is important to design Web content so that it is compatible with voice browser technology.

2. Current problems of voice browser technology

Voice browsers analyze the source code of Web pages (e.g., HTML (hypertext markup language) code) and use speech synthesis techniques to read it out. For example, the attributes associated with elements such as headings, lists, images, and tables are recognized, converted into sounds, and read out separately. For image files, the alternative text provided in the “alt” attribute is read out.

However, since the information is read out in the same order as it appears in the source content, it can be difficult for users to grasp information presented two-dimensionally, and it can take a long time to comprehend such information. For example, when the content shown in **Fig. 1** is read out by a voice browser, the text parts are strung together as shown by the speech bubble, making the title part and the hierarchical structure of the document unrecognizable. Also, current voice browsers do not have sufficient ability to analyze Japanese text and make lots of mistakes such as reading out *kanji* (characters of Chinese origin) with inappropriate readings.

3. Case study of Web accessibility with a voice browser

To understand the specific problems of Web design when the pages are accessed with voice browsers, we asked blind and sighted users to access the Web using voice browsers and tell us the problems they encountered. The sighted users were asked to listen to the voice browser while looking at the screen. The different categories of problems that were pointed out in this test are summarized in **Fig. 2**.

The problem most frequently pointed out by both groups was that of *kanji* being read out incorrectly. This problem depends on the dictionaries and lexical analysis functions built into the voice browsers.

† NTT Cyber Solutions Laboratories
Yokosuka-shi, 239-0847 Japan
E-mail: asano.yoko@lab.ntt.co.jp

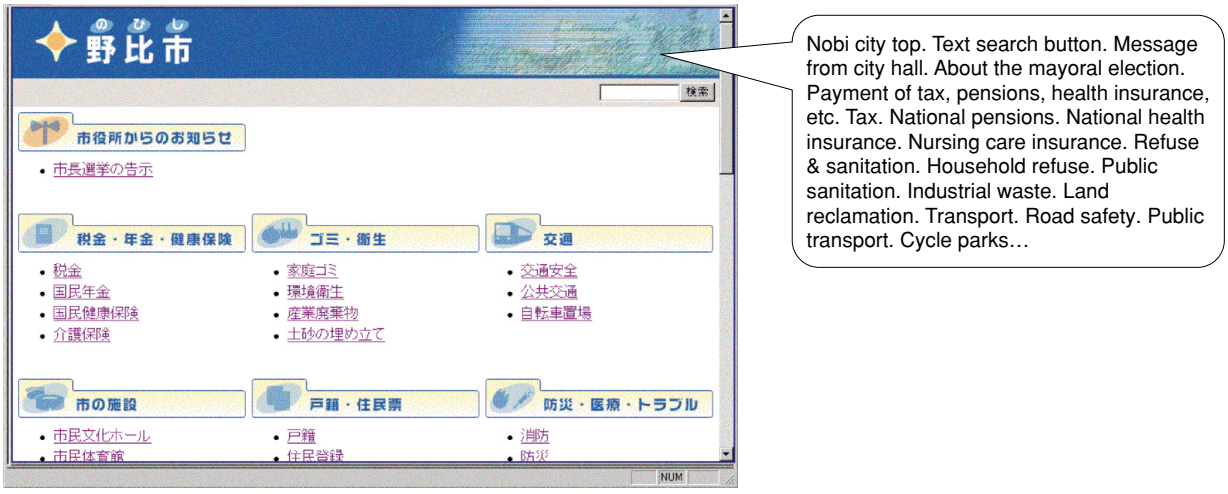

Fig. 1. Example of output from a voice browser.

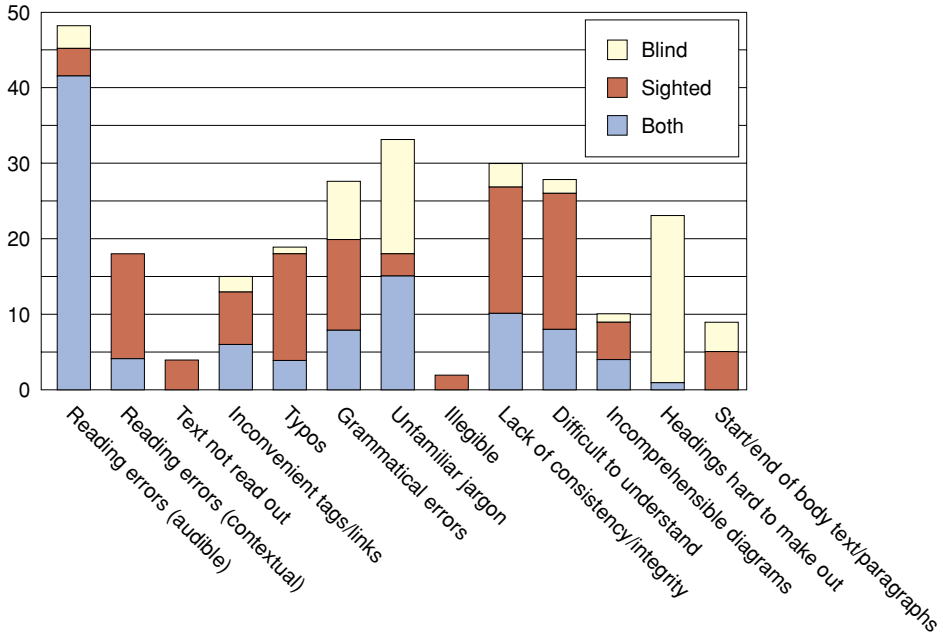

Fig. 2. Web accessibility problems pointed out by blind and sighted users.

On the other hand, the problems pointed out mostly by sighted users were content that was difficult to understand and content with a lack of consistency/integrity. Although these problems were pointed out less often by blind users, this probably does not mean that they understood the content in question, but rather that they were unable to grasp the overall flow of the context and content within the document.

Conversely, the problems that tended to be pointed out only by the blind users were the use of unfamiliar jargon and headings that were hard to make out. The problem of unfamiliar jargon is partly due to the fact that Japanese readers can understand unfamiliar terms by looking at their constituent *kanji*. But when text is presented purely in audible form, jargon that seldom crops up in ordinary conversation can often be difficult to understand. It can also be difficult to understand words with a short sound like 差 (*sa*: difference). We know this because the test subjects sometimes missed things and had to go back and lis-

ten to sections again.

As for the problem of headings being hard to make out, visual displays allow the grouping of text passages and the paragraph structure to be understood from visual cues such as margins, indents, and the size of the text. However, voice browsers simply read out the contents of a page from beginning to end without pausing, ignoring formatting information such as style sheets and text indents. As a result, it can be difficult to understand the structure of the information on a Web page, making headings difficult to distinguish from the body text. This sort of problem, which was only pointed out by blind users, tends to be overlooked by sighted users, so particular care must be taken when designing Web sites.

Other problems that were frequently pointed out included typos & transcription errors, content integrity, and HTML markup errors, but it should be possible to eliminate these by making thorough use of tools such as spellcheckers and HTML validators.

4. Design considerations for Web accessibility

Among the items pointed out in this evaluation, there were some problems that neither contravened any existing guidelines [1], [2] nor were capable of being detected by checking tools, but instead depended on human judgment. These problems arise because it is not possible to say specifically what is bad and what is unacceptable in such cases. For example, different voice browsers may have different ways of reading a given word in a given context, and certain symbols may be read out by some voice browsers but not by others. Such differences are difficult to confirm. They include problems of incorrect *kanji* readings, most of which will probably be eliminated in the future as the ability of voice browsers improves. However, it would be foolhardy to assume that all users are using advanced voice browsers, so it is important to take measures such as making an effort to use *kanji* expressions that are unlikely to be misread, avoiding the use of special characters or expressions, and using expressions that are unlikely to be misunderstood when considered objectively.

On the other hand, with regard to the difficulty of understanding how the information is structured [3], as pointed out by problems such as the headings being difficult to make out, it cannot be said that the problem is adequately covered by existing guidelines. Therefore, the following section presents a study specifically aimed at this sort of information structure design.

5. Design of speech navigation of information structure

When sighted users want to ascertain the structure of information on a Web page, before reading the specific content of the information, it seems that they first look for relationships between groups of information based on cues such as their positions and the presence of titles, and then form a framework of the information structure. Users also seem to grasp the overall amount of information in a page based on the amount of information presented and the positions of scroll bars. Section headings are also used to judge whether each section contains important information or can be skipped. However, since voice browsers read out the information in order from the beginning, it is difficult to grasp the structure of a page initially. We have therefore proposed a navigation system that starts by providing a spoken commentary on the information in a page when it is accessed by a voice browser. Specifically, this commentary might take the form shown in **Fig. 3**, where the following five types of description are added at the positions of the speech bubbles.

- The structure of the page is described at the beginning of the page.
- A page menu that links to each paragraph is added at the beginning of the page, allowing users to skip to the important parts.
- Numbers are added to the headings to help users understand their current position in the page.
- When there is a list containing several items, the total number of items in the list is given first.
- At the end of an itemized list or paragraph, the fact that the end has been reached is mentioned.

These descriptions contain information that sighted users will have already gathered from cues such as the visual layout, so the addition of further text information would be redundant in this case. We therefore employed a design where the information structure is only explained in speech form where it cannot be seen visually. There are a number of ways of achieving this, but for now we opted for the most commonly used technique of placing a transparent image in the document with the commentary text inserted into the “alt” attribute of this image from where it can be read out.

6. Evaluation of the speech navigation design

To evaluate what sort of comments and expressions are effective for helping users grasp the information

Fig. 3. Example of a spoken commentary on page structure.

structure of Web pages in a voice browser based on our proposed speech navigation method, we evaluated its performance alongside a Web site with the same content but without the added spoken commentary. In this evaluation, test subjects accessed the Web using voice information alone, without looking at the browser window, and we investigated points that were easy and difficult to understand when grasping the information content of Web pages.

We found that users' assessments were better overall for pages with a spoken commentary about the information structure. In sites with no commentary, users reported that they failed to understand the hierarchical structure and found it harder to distinguish item headings and lists in the content.

When individual commentaries were provided to explain the page structure at the start of each page, we found that the users did not take much notice of this description and failed to understand much of its content. In particular, when browsing by means of sound alone, it seems that it was not possible to grasp units such as "page", "body text", "menu", or "item", so there were also problems with the explanations. Similarly, when the voice browser announced that the end of a paragraph or list of items had been reached, users did not entirely understand what exactly had ended. Users also pointed out that the addition of this commentary complicated matters by making it impossible to distinguish between parts of the original content and the added commentary about the information structure.

On the other hand, when a page menu was added at the beginning of the page, even though that the users found it difficult to understand the term "page menu", they found it very useful for initially ascertaining the structure, size, and content of the whole page. The users also reported similar benefits from the initial description of the number of items in lists, which made things easier by enabling them to ascertain the overall amount of information in the page. Furthermore, the addition of item numbers to the item headings helped users recognize which parts of the text were headings, and this was found to be very useful as a means of following the order of the content and the current position within it.

The thing that gave users the most trouble was working out the hierarchical structure of the information and the positions of breaks within it. When a page menu was provided at the start of a page, containing links to the headings within it, we found that users mistook this page menu for a complete page and assumed that by following its links they were jumping to separate pages instead of to different parts of the same page.

These results show that the addition of speech navigation relating to the information structure is useful in that it allows users to imagine the information structure of a Web page before they have read it. In particular, it is very useful for providing commentaries to explain the quantity and sequence of information in a page. However, when information is acquired in speech form, it is liable to be grasped in

units that differ from the visual information structure, so it is important to be careful when using terms such as “page” to express information units, and pages must be designed so that the original content and the additional commentary can be easily distinguished from each other.

7. Conclusion

To check whether or not a Web page design is compatible with voice browsers, the best thing designers can do is to check their pages by listening to them being read out by a voice browser. Also, as our evaluation case studies have shown, having a Web page design evaluated with a real voice browser is likely to reveal many hitherto unnoticed problems, even with just a single user.

References

[1] Japanese Industrial Standards Committee, “JIS X8341-3: Guidelines for older persons and persons with disabilities—Information and communications equipment, software and services—Part 3: Web content,” Japanese Standards Association, Tokyo, Japan, 2004.
[2] M. Watanabe, A. Okano, and Y. Asano, “Universal Design Guidelines for Web Content,” NTT Technical Review, Vol. 3, No. 11, pp. 17-22, 2005.
[3] K. Fukuda, H. Takagi, J. Maeda, and C. Asakawa, “An Assistive Technology to Help Visually Impaired People Understand Web Page Structure,” Journal of the Human Interface Society, Vol. 5, No. 4, pp. 465-474, 2003.

Yoko Asano
Senior Research Engineer, Human Interaction Project, NTT Cyber Solutions Laboratories.
She received the B.E. degree in administration engineering from Keio University, Yokohama, Kanagawa in 1988 and joined NTT Human Interface Laboratories the same year. She moved to NTT Cyber Solutions Laboratories in 1999. Since then, she has been conducting research on human interfaces. She is a member of the Human Interface Society, the Japan Ergonomics Society, and the Information Processing Society of Japan.

Masahiro Watanabe
Research Engineer, Human Interaction Project, NTT Cyber Solutions Laboratories.
He received the B.E. and M.E. degrees in mechanical engineering in 1991 and 1993, respectively, and the Dr.E. degree in electronics, information, and communication engineering in 2003 from Waseda University, Tokyo. He joined NTT Basic Research Laboratories in 1993 and studied the human auditory system. He moved to NTT Cyber Solutions Laboratories in 1999. Since then, he has been studying human interfaces. He is a member of the Human Interface Society and the Acoustical Society of Japan.

Aya Okano
Research Engineer, Human Interaction Project, NTT Cyber Solutions Laboratories.
She received the B.S. degree in information science from the University of Tsukuba, Tsukuba, Ibaraki in 1998 and joined NTT Kanto Corporate Business Department the same year. She moved to NTT East Leased Circuit Service Center in 1999 and to NTT Cyber Solutions Laboratories in 2002. She is currently researching universal design.