Making Multimedia Content Accessible for Screen Reader Users

Hisashi Miyashita, Daisuke Sato
IBM Research, Tokyo Research Laboratory
1623-14, Shimo-tsuruma, Yamato-shi
Kanagawa 242-8502 Japan
{himi,daisuke}@jp.ibm.com

Hironobu Takagi, Chieko Asakawa
IBM Research, Tokyo Research Laboratory
1623-14, Shimo-tsuruma, Yamato-shi
Kanagawa 242-8502 Japan
{takagih,chie}@jp.ibm.com

ABSTRACT
Rich and multimedia content is increasing rapidly on the Web. It is very attractive for sighted people, but it brings severe problems to screen reader users. Once the audio starts playing, it becomes hard for blind users to listen to the screen reader because there is physically only one volume control that cannot control the separate audio streams. Though there are often software-controlled buttons to control the audio, they are often controllable only with a mouse and are not associated with alternative text. Because of the audio conflicts and inaccessible control buttons, the multimedia content is often inaccessible to blind users. In addition, the use of dynamically changing interactive user interfaces is also a critical issue, since existing screen readers cannot detect such dynamic content changes.

We developed an accessible Internet browser for multimedia to address these problems and offer multimedia content as an information resource for the blind. It is characterized by three major features. First, it allows users to control the audio, such as the volume, play/stop, pause, and even the speed. Second, a dynamically adaptable metadata function is added to simplify complicated multimedia pages and to track dynamic changes and effectively inform users about the changes. Third, an audio description function supports Internet movies with a text format described by the metadata. In this paper, after briefly discussing the existing accessibility problems of multimedia content, we describe our accessible Internet browser for multimedia.

Categories and Subject Descriptors

General Terms
Design, Human Factors.

Keywords
Accessibility, Multimedia control, DHTML, Flash, Streaming Media.

Permission to make digital or hard copies of all or part 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. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM Reference Format

ACM Reference Format

1. INTRODUCTION
The Web is extending to a wider range of applications in new websites, and multimedia content is becoming increasingly popular for richer user experiences. Many prominent websites such as YouTube (http://www.youtube.com) and Yahoo! Video (http://movies.yahoo.com) distribute multimedia content and are accessed by enormous numbers of users. Typical multimedia content has a fancy GUI and various streaming media. However, such content poses significant problems for blind users with existing screen readers.

There are two major problems. First, blind users need to hear the speech from their screen readers to navigate in a page, but the media’s sound is often masking it, so they cannot hear the screen reader and lose control of the page. This situation occurs because there is only one physical volume control and it is not possible to control each sound separately. Even when an application provides its own controls, these controls are sometimes only for a mouse and have no alternative text. The result is that blind users cannot deal with the multimedia audio content. Second, rich multimedia content pages are visually active and some parts of the page are often dynamically changing without any user interaction. The changes are visually obvious, but there are no non-visual cues. Existing screen reader technology cannot detect such dynamically changing content and thus, no information about the updates is given to the users.

2. MULTIMEDIA ACCESSIBLE BROWSER
We developed an accessible Internet browser for multimedia to solve these problems and to offer multimedia content as an information resource for the blind. The multimedia accessible browser (Figure 1) is characterized by three features: 1) Non-visual multimedia audio controls, 2) Browser integration of audio descriptions, and 3) An alternative user interfaces using external metadata. It can work as a self-talking browser or work with a screen reader for blind users to use their favorite screen reader.

2.1 Non-visual Multimedia Audio Controls
Independent of any audio controls provided by content developers, our multimedia browser provides audio control buttons that are assigned to shortcut keys and which can be used at any time. When users hear sound from a multimedia file, they can use these shortcut keys. Ctrl-P will pause or play, Ctrl-J and Ctrl-K will increase and lower the volume, respectively, Ctrl-S will stop or rewind, and Ctrl-M will mute or unmute the sound. In addition, Ctrl-Up/Down increases/decreases the playback speed. To support these functions, the browser searches for sound and media objects in the webpages and controls them directly to supply an effective
user interface. This solves the problems so blind users can control the multimedia audio content.

This means blind users can use Ctrl-P to pause/resume or Ctrl-K and Ctrl-J to adjust the volume for any multimedia content. When blind users want to hear the screen reader for navigation, they can press Ctrl-J to decrease the volume of the media in the page without affecting the volume of the screen reader. Alternately, they can use Ctrl-P to pause the media. In either way, they can immediately control the media without searching in the content. Since blind users require speech to navigate in the content, this immediate media control is indispensable to avoid audio interference with the guiding speech from the talking browser.

The variable speed control feature is helpful for many blind users, since they rely heavily on audio information and have developed the ability to listen to high-speed voices [3]. If they feel the speed is too slow, they can increase the speed with Ctrl-Up. Alternately, to listen more carefully after missing some part of an audio, they can rewind the media with Ctrl-Left and then replay that part more slowly using Ctrl-Down.

2.2 Audio Description Integration
Some TV programs provide audio captions for blind people to help them better understand what is on the screen. However, such additional information is rarely used with Internet video. Our system supports adding audio captions in a text format that can be included in the metadata described in the next section. In this paper, we call these audio captions the audio description. The advantage of an audio description is that it is easier to provide in comparison to an audio caption recorded by a person.

The timeline for each description is stored to synchronize the audio description with the video. The browser tracks the timeline of the active media, and reads each description at the specified time. It also tracks the volume of the media to allow users to easily listen to the audio description along with the multimedia sound. It can automatically adjust the volume of the audio separately from the media if it is too loud. Of course, users can manually and separately control the volumes of the audio caption and media. The speech rate for the audio description is also controllable. Even when users change the speed of the media, the audio description remains synchronized with it.

2.3 Alternative User Interface with Metadata
Our browser can provide an alternative user interface for multimedia Web content by making use of external metadata. Since multimedia content is often too complicated and changes dynamically, blind users get lost. In addition, most pages with multimedia content do not provide any alternative text for images, even though there are many images in those pages. Also, they often support only mouse operations for interactions [1]. As a result, blind users cannot obtain any information with existing screen readers nor control the pages with the keyboard.

To solve this critical situation, our browser can use external metadata with inaccessible multimedia content. The metadata describes multimedia content such as Dynamic HTML and Adobe Flash. By using this feature, the content is simplified and has a more easily accessible user interface for blind users. The metadata is described in XML and it can supply essential information such as the logical structure of an alternative user interface, alternative text, the roles of the widgets, and data about the events in important dynamic changes. Our browser can attach such information to dynamic Web content even while the users are navigating in the changing content. Although there exist many studies that apply external metadata to static HTML-based pages and transcode them into accessible content [2], our work is the first attempt to apply external metadata to dynamic Web content. Even when the content is changed by navigation, our browser handles the changes appropriately by adding essential information to the changed parts and by reporting on the updates as needed.

3. CONCLUSIONS
After introducing existing accessibility problems faced by blind users in accessing the multimedia content, we gave an overview of our accessible Internet browser for multimedia. It has three primary features: (1) Non-visual multimedia audio control, (2) An audio description mechanism, and (3) Dynamically generated alternative interfaces. If accessible features for multimedia content are available, then blind users will be able to control that content. Our future direction involves open-sourcing the technology and providing ways to describe the metadata effectively. As accessible multimedia content increases, users will be able to benefit.

4. ACKNOWLEDGEMENT
We thank Bob Regan and Andrew Kirkpatrick at Adobe Systems for their suggestions and comments while developing this system.

5. REFERENCES