A Hyper Presentation System for Mobile Web Browsing

5T-02

Zhiyong *Peng*+ Michal *Trebacz*++ Junichi *Yamazaki*+
HP laboratories *Japan*+, University of Waterloo *Canada*++

1 Introduction

Documents that are distributed over the World Wide Web are typically specified in hypertext markup language (HTML). Users can browse these documents using a browser. To date, most browsers assume that the programs are running on high-performance hardware, with at least 16 megabytes of memory, a high speed modem, a relatively large display, and so on. However, the same documents presented on low-performance hardware such as present in handheld computers with small display areas perform poorly in mobile environment. For example, the small size of the display area necessitates frequent scrolling of the page to access all of the data on a web page.

In the mobile environment, sound data is easy to be accessed. In systems that use conventional browsers, a sound tract consisting of a human speaking ("narration speaking") is usually stored as sampled audio data consisting of amplitude and phase information as a function of time. Accordingly, keyword searches of files on the internet performed by accessing a web search engine cannot select files based on the content of human speech rendered as sampled sound. In conventional systems, the user must provide some form of written abstract of the sampled sound content in the document to assure that the document will be found during keyword searches. Providing such abstracts is time consuming and increases the storage and bandwidth needed to process the document.

In such cases, web users cannot do a complete search in order to find out the contents of a narration, since possible methods such as including a previously compiled abstract of the narration in an HTML document or attaching the abstract to an HTML document do not search the narration that is not specified in the abstract. Sampled sound files can present additional problems when played back on hardware platforms having limited computational or storage capabilities. On such systems, the sound track may be out of synchronization with the non-sound portion of the display. This is particularly true in systems having limited bandwidth for the transfer of the various files between the server and the browser.

In order to solve the above problems, we proposed a hyper presentation system, which is based on the observation that a typical page viewed on a browser includes both graphic and textual material. We provide a modification to the hypertext markup language that allows the typical web page to be easily converted to a sequence of "slides" representing the graphical information with the textual information being presented as a sound track by Text-To-Speech engine. The slides are switched in accordance with the flow of the spoken text. In this manner, even a large web page can be viewed by a handheld computer with a small display area. Furthermore, the narration contents can be easily found out by the web search engine because they are stored in form of texts.

2 Hyper Presentation System

A hyper presentation system is characterized by having a browser that downloads source files according to a suitable protocol such as hyper text transfer protocol (HTTP), and a processor having a display and speaker system. Source files are basically text files specified in the modified HTML language. In the following description, a language that adds the additional tags discussed below to HTML will be referred to as hyper presentation markup language (HPML). Files written according to the HPML specification will be referred to as HPML files.

The specification methods for hyper presentation markup language include a slide showing tag that controls the slide display and a narration tag that controls the narration speaking of the character strings. In addition, a pause tag is also defined for stopping interpretation of script for a specified period of time.

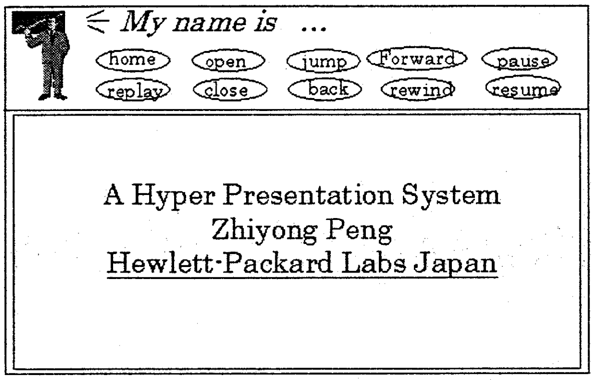
A slide showing tag is composed of a slide start element and a slide end element. The slide start element defines the content of the slide and places the image on the display. The slide end element erases the slide image on the display. The layout of the slide content is defined by HTML tags, such as <center>, <p>, <table>, <ul>, <img>. Slide showing tags can be specified in a nested structure so that slides can be pushed onto a stack and popped back to the display during a presentation.

Similarly, narration speaking tags preferably include a narration start element and a narration end element. The narration start element is typically placed after the corresponding slide start element. The narration start element executes a process that converts character strings, to signals representing narration and outputs them from the speaker. The character string can be part of the tag or specified by a file included in the tag. The narration end element halts the output of narration speaking.

The processor fetches the source files, shows slides on the display according to the slide showing tags and displays other information written in the conventional HTML language. The processor converts character strings specified by the narration speaking tags into

```
<HPML>
<SLIDE>
<P>
<CENTER>
<H1>A Hyper Presentation System</H1>
<P>
<H2>Zhiyong Peng</H2>
<P>
<H2><A HREF="hpl.hpml">
Hewlett-Packard Labs Japan</A></H2>
</CENTER>
<PAUSE T= 3>
<P>
<NARRATION>
My name is Zhiyong Peng. I am from
Hewlett-Packard Labs Japan, and will
introduce a hyper presentation system.
</NARRATION>
</SLIDE>
.....
```

HPML Player



My name is ...

A Hyper Presentation System
Zhiyong Peng
Hewlett-Packard Labs Japan

audio signals and outputs them to speakers based on the narration speaking tags.

Slide showing and narration speaking are synchronized according to their sequence and nested structure. The links within slides shown on the display and links within the sound information output from the speaker can be used to trigger another presentation by pointing to a link in a slide.

3 A Prototype Implementation

We implemented a HPML player which can be used to browse hyper presentation contents on Web. As shown in above figure, the slide defined by HPML on the left is viewed by the HPML player on the right. The player is designed with three areas. The first area is used to emulate a number of buttons to control hyper presentations. The second area is used to display a "slide". The third area displays the narration text from which the sound track is being generated. Hot spots are included in the text in the slide shown or in the sound information output from the speaker. The user can jump to a URL by clicking its hot spot with a mouse pointer. The hot spot in the sound can be produced with a beep that lets the user know that it is a hot spot. A user whose attention is drawn by this beep can press the JUMP button to jump to the URL linked to it.

4 Conclusion

In real life, people usually get information by reading documents, listening radio, watching TV programs and listening presentations. These years, a lot of technologies have been developed to support information broadcasting on web. For example, HTML was designed to define documents on the web. There are VoiceXML[2] and TVML[3] which are used to specify voice and TV programs on the web, respectively. SMIL[1] was introduced for multimedia presentations. Our solution is like SMIL but very simpler than it. We define layout of slides by HTML tags and extend HTML with new tags for specifying synchronization between slide showing and narration speaking. Users having HTML knowledge can easily create presentations by HPML. Because the HPML files convert a large page into a sequence of small slides and don't include media having the large amount of data such as audio and video, they can be easily browsed by hardware platforms having limited computational or storage capabilities in the mobile web environment.

Reference

[1] SMIL 1.0 Specification: http://www.w3.org/TR/REC-smil/

[2] VoiceXML 1.0 http://www.w3.org/TR/voicexml

[3] TVML http://www.strl.nhk.or.jp/TVML/