MPML: A Multimodal Presentation Markup Language with Character Agent Control Functions

Takayuki Tsutsui, Santi Saeyor and Mitsuru Ishizuka
Dept. of Information and Communication Eng., School of Engineering,
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 JAPAN
{tsutsui,santi,ishizuka}@miv.t.u-tokyo.ac.jp

Abstract: As a new style of effective information presentations and a new multimodal information content production on the World Wide Web (WWW), multimodal presentation using interactive life-like agents with verbal conversation capability appears to be very attractive and important. For this purpose, we have developed Multimodal Presentation Markup Language (MPML), which allows many users to write attractive multimodal presentations easily. MPML is a markup language conformed to Extensible Markup Language (XML). It supports functions for controlling verbal presentation and agent behavior. In this paper, we present the specification, related tools, and application of MPML when used as a tool for composing multimodal presentations on the WWW.

Introduction

Providing attractive and effective information to all ranges of audiences becomes an important matter to the information providers. We consider that using character agent to provide multimodal presentation, as a new form of presentation, is attractive and significant. Currently available presentation tools provide explanation screens and displaying features for human presenter to manipulate and deliver the presentation by voice and actions. This is comfortable for the audiences to perceive compared to the plain document. In this paper, we propose a multimodal presentation by character agent instead of human presenter. The effort has been made to provide the character agent with various features so that it can deliver the presentation without intervention of human presenter, which is a desirable feature for contents on the WWW.

Nowadays, the developments in character agent system and voice recognition/synthesis are very sophisticated so that such a presentation can be made practical. However, it is subtle and tedious task to make content like that because of the specific features including script language in each system. In order to promote the use of such content, it is necessary to innovate a script language that works together with HTML and simply enough for the content builders to incorporate into their pages.

Presentation Agent

At present, not only text and graphics can be used in WWW pages, we may compose multimedia presentation by putting animation, music and voices on the pages. Such scenario is emerging quickly. The content makers can create their presentation and provide it on the WWW so that everyone can access the presentation anytime. (Fig. 1 (b))

Even that seems to be quite fascinating, it is only one-way communication. The users may use mouse to jump from page to page or close the window, which are bound for page hopping and session ending. Moreover, such fashion is different from the presentation performed by human as shown in (Fig. 1(a)). The audiences cannot feedback their feeling to the content makers.

We considered implementation of a character agent system, which enables us to provide fascinating multimodal information content and presentation on the WWW. Some examples of currently available multimodal anthropomorphic agent interface are the TOSBURG II by Toshiba (Takebayashi 94), the system at Sony CSL (Nagao 94) and the system at Electrotechnical Laboratory (Hasegawa 97). At the moment, there are many research works on

automatic presentation such as Virtual Human Presenter at University of Pennsylvania, and WebPersona, which have WWW capability at The German Research Center for Artificial Intelligence (DFKI).

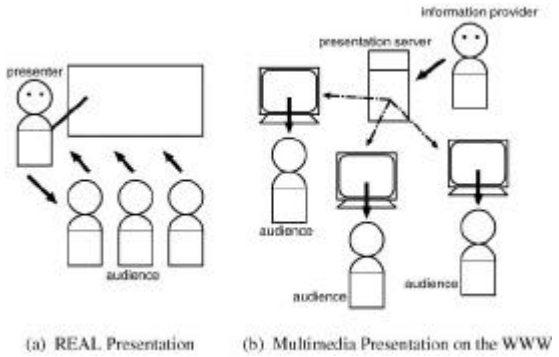

Figure 1: Styles of Presentation

Character Agent based Presentation on the WWW

Using character agents to present contents on the WWW is a promising way to make the contents attractive and promote widespread use of multimodal contents as shown in (Fig. 2).
In order to make use of presentation on the WWW by character agents, an important issue is that we should have a scripting language, which is easy to compose and does not depend on each character agent system. In this paper we have designed and developed Multimodal Presentation Markup Language (MPML) as the first step to achieve the objectives described above.

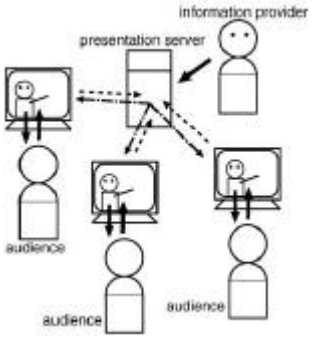

Figure 2: Multimodal Presentation by Character Agent on the WWW

Features of Multimodal Presentation Markup Language MPML

MPML is a markup language, which is designed and developed to facilitate multimodal presentation by character agents. It has the following features:

?? **Platform Independency**: The content builders usually need to take audiences' OS, browsers and resources into account when providing presentation on the WWW. MPML is independent to browsers or systems. Moreover, it is designed so that the contents written in MPML can be played on wide variety of tools or players.
?? **Simplicity**: MPML conforms to XML (Extensible Markup Language) specification. At the present, MPML version 1.0 implements 19 tags. For those who can write HTML scripts to build web pages, they will find that writing multimodal presentation by character agents in MPML is quite simple.
?? **Media Synchronization**: Synchronization of medias such as voices, images and gestures is necessary to create an attractive presentation. On this purpose, W3C announced SMIL (Synchronized Multimedia Integration

Language) (see SMIL), which is a language for controlling complex media data on the WWW in 1998. MPML implements media synchronization based on SMIL specification.

?? **Controls of Character Agents**: MPML supports action controls of character agents such as greeting, pointing and explaining. Furthermore, the expression controls such as smiling and puzzled are also incorporated.

?? **Controls of Interactive Presentation**: MPML also supports the use of hyperlinks. When using with voice recognition engine, it can conduct the interaction between the audience and the character agent via voice commands, which serves well as navigation along the presentation.

Figure 3: MPML structure tree

Specification of MPML

This section is devoted to explain the specification of MPML. The tree diagram that represents the structure of MPML document is shown in (Fig. 3). The mark "?" indicates that the tag can be omitted or used at most 1 time. The mark "*" indicates that the tag can be used any arbitrary times. The "#PCDATA" in the tree diagram represents text data. The root of all elements is the tag <mpml>, which has "id" attribute. The "id" attribute is utilized to facilitate identification of tags. Most of the tags can be assigned IDs.

Document Headers

Content builders can provide information about the presentation and layout in MPML document using the area cast by the <head>…</head>. Meta data can be provided by using tag <rec> and layout information can be provided using tag <layout>.

?? **Meta Data**: Content builders can write general information about the presentation using <meta> or <abst> within tag <rec>. Tag element <meta> is an empty content tag in which information can be put as its attribute. Tag element <abst> is a content-defined tag. The contents of the tag control the layout of the presentation.

?? **Layout**: The contents of tag element <layout> are the information about layout of the presentation. The contents can be arbitrary but MPML has its default layout style. The sub element can be <root-layout> or <region>. Tag element <root-layout> defines the characteristic of the root window of the presentation. Tag element <region> defines layout information for points or rectangular regions. The content builders can use one tag <region> for one region.

Document Body

The document body cast by <body>…</body> contains the contents of the presentation. By default, the tag element <body> contains <seq>. If there is nothing specified, the actions will be sequential.

?? **Agent Selection**: Tag element <agent> is used to select the character agent that performs the presentation. Tag element <move>, <speak> and <play> will refer to the agent given in tag element <agent>. The content builders can use multiple agents to perform the presentation by using <agent> to initiate agents with corresponding IDs.

?? **Agent Movement**: The content builders can move character agents using tag element <move>. The agents can be moved to defined regions or points or to specified coordinates.

The content of tag element <speak> is text sentences. The players send this information to voice synthesizer engine of the character agents to make them speak. Moreover, tag element <play> can be used to play actions of character agents. MPML is capable of playing basic actions such as greeting, pointing to selected regions, and doing some actions at the same time. The attributes of each tag element are listed in (Tab. 1).

Table 1: Tag elements for agent behavior description

Tag Element	Attribute	Function
move	id	Identification
	agent	Specify id of <agent> to be moved
	region	Specify id of destination
	location	Specify coordinates of destination
	stand	Specify standing point for destination
	speed	Specify moving speed
speak	id	Identification
	agent	Specify id of <agent> to speak
	lang	Specify language to speak
	voice-type	Specify type of the voice
	speed	Specify speaking speed
	begin	Specify the time to start speaking
	end	Specify the time to stop speaking
	dur	Specify speaking duration
	alt	Specify using of message display when voice is not supported.
play	id	Identification
	agent	Specify agent to do actions
	act	Specify action content
	parts	Specify the parts to do actions
	object	Specify object id to do actions
	object-loc	Specify coordinates of the action
	degree	Specify level of actions
	speed	Specify speed of doing actions.
	begin	Specify the time to start actions
	end	Specify the time to stop actions
	dur	Specify duration of action
	track	Select to enable/disable tracking
	point-gesture	Specify hand actions when doing actions

Media Synchronization

The contents of tag element <par> will be played in parallel regardless of the orders in the list. For example, the action model shown in (Fig. 4(a)), the character agent will start speaking 2 seconds after initiated greeting.

The contents of tag element <seq> will be played sequentially according to the order written in the list. For example, the action model shown in (Fig. 4(b)), the character agent will start speaking 2 seconds after the greeting action is done.

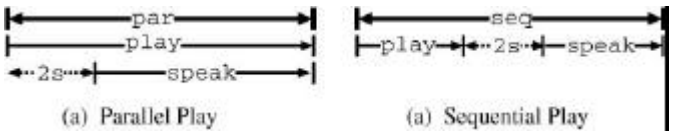

Figure 4: Models of synchronized play in MPML

Hyperlink/Presentation Controls

The content builders can control the presentation according to mouse operations or voice input by using tag element <a>. With this tag, the content builders can stop, restart the presentation, and jump among portions.

Tag element <a> can have the following attributes:

id, title, href, show, mode, begin,
end, dur, region, listen, lang,
key, confidence

The attribute mode is used to determine the control when interaction occurs. For example <a mode="jump" href="link1"> the presentation will jump to the specified link. The attribute href is similar to that of HTML, which specifies a link. The attribute show is used to control the flow of presentation when jumping to another link happened. The attribute key is used to specify the input voice commands. Together with this attribute, selection mark " | ", option marks "[ ]", shortcut mark "…", priority marks "( )" are used. For example, key="…[say] (hello | hi)" the input can be either "say hello", "please hello", or "Hi". The attribute confidence is used to specify the reliability level of recognized voice commands. The attribute listen lets the character agent to wait for an input voice command, where the input can be controlled by begin, end, and dur. The attribute region specifies the region on which the specified actions take place when the mouse is clicked.

Moreover, the tag element <anchor> can be used in 2 ways as follows:

?? To specify a terminal anchor on a point of time of media object or agent action.
?? To specify an anchor on point of space of media object or agent action.

The content builders can jump into the middle of media object or agent's dialogue specified by <anchor> using the tag <a>. For example, if the presentation jump to the point id called "anc1" the talk will start right from that point.

```
<speak>
   This part will not be read
   If jump-started from "anc1".
   <anchor id="anc1" />
   Read from here!
</speak>
```

Presentation of alternative contents

Tag element <switch> is designed to use when the contents do not match the capability of the player. MPML enables using of multiple alternatives. The content builders can provide the contents in a variety of formats sorted by the preferable forms. Usually, text content comes the last order since text capability is likely to be the most basic feature of any players.

Sample Script

A simple MPML document can be written as shown below:

```
<mpml>
    <head>
        <layout>
            <root-layout id="root" width="800" height="600" />
            <region id="spot1" location="500, 300" />
            <region id="spot2" location="20%, 50% />
        </layout>
    </head>
    <body>
        <ref region="root" src=http://www.miv.t.u-tokyo.ac.jp/" />
        <agent region="spot1" />
        <par>
            <play act="point" region="spot2" />
```

```
            <speak>
                This is MPML Home Page!
            </speak>
        </par>
    </body>
</mpml>
```

With the above script the presentation will be like following:
The background of the presentation will be the web page, which the URL is specified in <ref>. The default agent character will appear from the area specified by spot1. The agent will point to the area specified by spot2 and speak the content bounded by <speak> and </speak>.

Comparison with Other Markup Languages

The comparison of MPML with other markup languages (SMIL and HTML) is shown in (Tab. 2).

Table 2: Comparison of MPML with other WWW markup languages.

Scripting Function	MPML	SMIL	HTML
Web publication	Possible	Possible	Possible
Link to other URLs	Possible	Possible	Possible
Data description	Has standard form	Self defined	Basicly impossible
Layout description	Possible	Possible	Basicly impossible
Media Synchronization	Minimum features	Full features	Impossible
Agent's action description	Possible	Impossible	Impossible
Mouse Control	Possible	Possible	Possible
Voice Control	Possible	Impossible	Impossible
Text to speech	Possible	Impossible	Impossible
Current users	Very little	Few	Remarkably large
Tools	Few	About 10	A great number
Number of tags	About 20	About 20	About 80

Even all these markup languages are designed for Web publication, there are some differences. For example, since SMIL is designed mainly for media synchronization, the description of layout and timing for playing the media are strengthened in its specification. On the other side, since MPML is designed mainly for simplicity in character agent based multimodal presentation content composing, it incorporates only minimum media synchronization and layout features sufficient to perform presentation. Furthermore, due to the need of speech dialogue features, it has to incorporate voice commands and TTS (Text-To-Speech) capability.

Concluding Remarks

The MPML is a script language that facilitates the making and distributing of multimodal contents with character presenter is proposed. MPML is conforms to XML specification. At the same time, it supports media synchronization with character agents' actions and voice commands that conforms to SMIL specification. The content builders can use MPML to create multimodal presentation contents on the WWW simply by scripting with the small set of MPML tags. At the moment, only basic interactive functions, which are sufficient to the presentation aspect, are available. However the bi-directional communication between the contents and the audiences should be studied more and incorporated into the MPML specification in order to expand the application areas.

References

[Takebayashi 94] Takebayashi Youichi (1994): Free Speech Dialogue System TOSBURG II – Toward the Realization of User-centered Multimodal Interface, *Journal of Electronics, Information and Communication Engineers*, Vol. J77-D-II, No. 8, pp. 1417-1428

[Nagao 94] Nagao, K. and Takeuchi, A. (1994): Speech Dialogue with Facial Displays: Multimodal Human Computer Conversation, Proc. 32nd *Annual Conf. of ASSOC of Computational Linguistics*, pp. 102-109

[Hasegawa 97] Hasegawa, O. and Sakaue, K. (1997): A CG Tool for Constructing Anthropomorphic Interface Agents, Proc. *IJCAI-97 Workshop on Animated Interface Agents: Making Them Intelligent*, Nagoya, Japan, pp. 23-26

[Noma 97] Noma, T. and Badler, N. (1997): A Virtual Human Presenter, *IJCAI-97 Workshop on Animated Interface Agents: Making Them Intelligent*, Nagoya, Japan, pp. 45-51

[Andre 98] Andre, E., Rist, T. and J. Muller (1998): WebPersona: A Life-Like Presentation Agent for the World Wide Web, *Knowledge-Based Systems*, Vol. II, pp. 25-36

[XML] XML HomePage: http://www.w3.org/XML

[SMIL] SMIL HomePage: http://www.w3.org/AudioVideo/

[MPML] MPML HomePage: http://www.miv.t.u-tokyo.ac.jp/MPML/mpml.html

[Clark 98] Clark, D. and Stuple, S. J. (eds.) (1998): Developing for Microsoft Agent, *Microsoft Press*