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

Mitsuru ISHIZUKA, Takayuki TSUTSUI, Santi SAEYOR, Hiroshi DOHI, Yuan ZONG, and Helmut PRENDINGER

Dept. of Information and Communication Engineering, The University of Tokyo.

Abstract

As a new style of effective information presentation and a new multimodal information content production tool for the 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 MPML (Multimodal Presentation Markup Language), which allows users to write attractive multimodal presentations easily. MPML is a markup language conformed to XML (Extensible Markup Language). It supports functions for controlling verbal presentation and scripting agent behaviors. In this paper, we present the specification, related tools, and applications of MPML when used as a tool for composing multimodal presentations on WWW.

1. Introduction

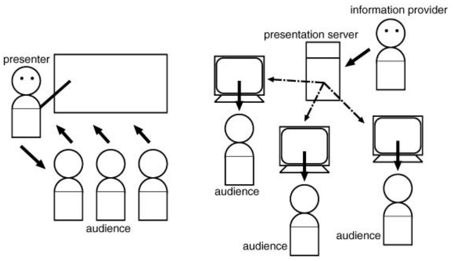
Providing attractive and effective information to different types of audiences becomes an important issue for the information providers. We believe that using character agents to provide multimodal presentation, as a new form of presentation, is attractive and significant. Currently available presentation tools provide explanation screens and displaying features for a 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 a character agent instead of a 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 a human presenter, which is a desirable feature for contents on WWW.

Nowadays, the developments in character agent systems and voice recognition/synthesis are very sophisticated so that such a presentation can be made practical. However, it is a subtle and tedious task to create content in this way 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 is simple enough for the content builders to incorporate into their pages.

2. Presentation Agent

The content authors can create their presentation and provide it on the WWW so that everyone can access the presentation anytime. (Fig. 1 (b))

Even though this seems to be quite fascinating, it is only 'one way communication'. Moreover, this presentation style is different from the presentation performed by human as shown in Fig. 1(a), where audiences can give feedback to the content authors.



(a) REAL Presentation (b) Multimedia Presentation on the WWW

Fig. 1 Styles of Presentation

Some examples of currently available multimodal anthropomorphic agent interfaces are TOSBURG II by Toshiba [1], the system at Sony CSL [2] and the system at Electrotechnical Laboratory [3]. At the moment, there are many research works on automatic presentation such as Virtual Human Presenter at the University of Pennsylvania, and WebPersona at The German Research Center for Artificial Intelligence (DFKI), which has WWW capability.

2.1 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 use and does not depend on each character agent system. In this paper we have designed and developed a Multimodal Presentation Markup Language (MPML) as the first step to achieve the objectives described above.

3. Features of the 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:

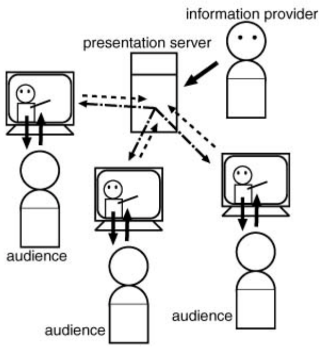


Fig. 2 Multimodal Presentation by Character Agent on the WWW

- **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 of 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.
- **Media Synchronization**: Synchronization of medias such as voices, images and gestures is necessary to create an attractive presentation. For this purpose, in 1998, W3C announced SMIL [7], which is a language for controlling complex media data on the WWW. MPML implements media synchronization based on the SMIL specification.
- **Control of Character Agents**: MPML supports action controls of character agents such as greeting, pointing and explaining. Furthermore, expression controls such as 'smiling' and 'being puzzled' are also incorporated.
- **Control of Interactive Presentation**: MPML also supports the use of hyperlinks. When used 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.

4. Specification of MPML

This section is devoted to explain the specification of MPML. The tree diagram that represents the structure of MPML 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 arbitrary times. "#PCDATA" in the tree diagram represents text data. The root of all elements is the tag <mpml>, which has an "id" attribute. The "id" attribute is utilized to facilitate identification of tags. Most of the tags can be assigned IDs.

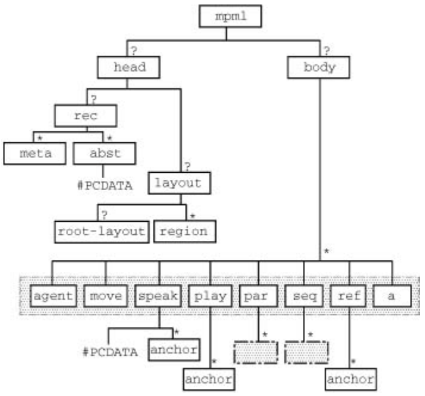


Fig. 3 MPML structure tree

4.1 Document Headers

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

- **Meta Data**: Content builders can write general information about the presentation using <meta> or <abst> within tag <rec>. The 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 content of the tag controls the layout of the presentation.
- **Layout**: The content of tag element <layout> is the information about the layout of the presentation. 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**: The tag element <agent> is used to select the character agent that performs the presentation. The tag element <move>, <speak> and <play> 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. This information is sent to voice synthesizer engine to make agents speak. Moreover, the 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 Table 1.

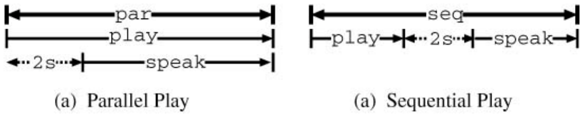
Table 1 Tag elements for agent behavior description

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, in the action model shown in Fig. 4(b), the character agent will start speaking 2 seconds after the greeting action is finished.



(a) Parallel Play (a) Sequential Play

Fig. 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 within specified regions.

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, with <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 wait for an input voice command, where the input can be controlled by begin, end, and dur. The attribute ‘region’ specifies the region in which the specified actions take place when the mouse is clicked.

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

- To specify a terminal anchor that determines the stopping time point of a media object or agent action.
- To specify an anchor that determines the coordinates of a media object or agent action.

The content builders can jump into the middle of media object or the 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

The tag element <switch> is designed to be used when the content does not match the capability of the player. MPML enables the use of multiple alternatives. The content builders can provide the contents in a variety of formats sorted by preferable forms.

5. Composing Presentation in MPML and Experiment Software

The aim of this work is to enable composition, distribution, and viewing character agent based multimodal presentation made by MPML on the WWW. The audiences can view the presentation in the way shown in Fig. 5. In the following we explain each related tool.

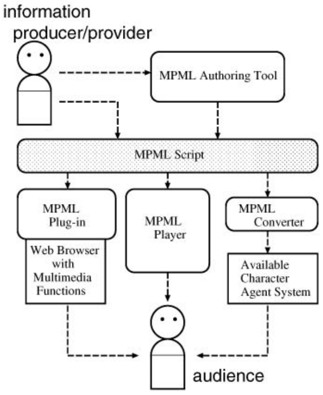


Fig. 5 MPML-related software tools

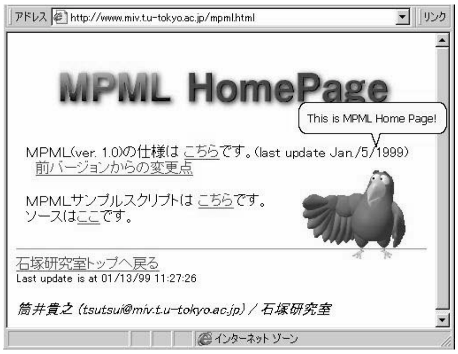
MPML Player

This is the most important tool that loads MPML script and plays the contents. MPML Player implements a set of character agents and media players to carry out the presentation. In order to implement the MPML Player, we need the following technologies:

- Character agent with control functions
- Speech Recognition/Synthesis engine that supports English/Japanese
- Media player that have synchronization function
- A set of actions that is independent of any systems
- Web tools

It is obvious that MPML Player has to deal with many complex components. At the moment, we have tested and developed MPML using converters in order to carry out the presentation by existing tools.

Currently, we have two converters for MPML script. The first one is applicable to Microsoft Agent System [9]. The converter called MPML2MSAScript converts MPML script into VBScript, which is capable of controlling Microsoft Agent. Due to the limitation of the system, this combination cannot support all MPML features but the following features are possible:


アドレス http://www.miv.t.u-tokyo.ac.jp/mpml.html リンク

MPML HomePage

This is MPML Home Page!

MPML(ver. 1.0)の仕様は こちらです。(last update Jan./5/1999)
前バージョンからの変更点

MPMLサンプルスクリプトは こちらです。
ソースはここです。

石塚研究室トップへ戻る
Last update is at 01/13/99 11:27:26

筒井貴之 (tsutsui@miv.t.u-tokyo.ac.jp) / 石塚研究室

インターネット ゾーン

Fig. 6 Snapshot of presentation generated by MPML2MSAScript

- Seamless synchronized media playing
- Moving character to arbitrary points
- English speech recognition and synthesis
- Controlling the presentation by mouse or voice commands
- Displaying linked web pages

The screen when running the VBScript version of MPML script, which was converted by MPML2MSAScript, is shown in Fig. 6.

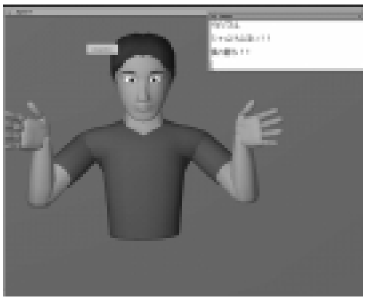


Fig. 7 Snapshot of presentation generated by MPML2Ashow

Another converter is the converter that converts MPML script into control codes of Ashow [3], a CG tool with precise hand and finger motions. Fig. 7 shows the result of using MPML and its converter called MPML2Ashow. Tough the motions in the Ashow system are different from the actions defined in

MPML, the converter uses an action mapping technique to generate the control code for Ashow. By using the same technique, we can also program converters for other character agent systems.

6. Evaluation

The comparison of MPML with other markup languages (SMIL and HTML) is shown in Table 2. Even though all these markup languages are designed for Web publication, there exist 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.

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

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.

6.2 Evaluation and Discussion

Obviously, using MPML to compose a character agent based multimodal presentation is a lot easier than using other markup or script languages. For example, Microsoft Agent, the character agent system from Microsoft, can be programmed in VBScript but due to the complexities and structures of the programming language, it is still difficult for general users to make multimodal presentation script with this. MicrosoftAgent was released in 1997 but did not gain wide popularity among common users, which might be because of difficulties in programming.

MPML is designed especially for this purpose so it has simple specification but covers most significant features. It hides overheads and complexities in scripting from general users. As we can see in the case of Microsoft Agent, the content composers write presentation script in MPML, and then use a converter like MPML2MSAScript to convert the script into native script language for the IE browser. MPML simplifies the content composing to a level that is easy compared to other languages.

7. Concluding Remarks

MPML is a script language that facilitates the production and distribution of multimodal contents with character presenter. MPML 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 for the presentation aspect, are available. However, the bi-directional communication between the content presenter and the audiences should be studied more and incorporated into the MPML specification in order to expand its use to more application areas.

8. References

[1] Takebayashi Youichi: 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, 1994

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

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

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

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

[6] XML HomePage: http://www.w3.org/TR/REC-xml/

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

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

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