Visual Authoring Tool for Presentation Agent based on Multimodal Presentation Markup Language

Santi Saeyor, He Binda 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
{santi,nani,ishizuka}@miv.t.u-tokyo.ac.jp

Abstract: This paper proposes a visual authoring tool for presentation agent system on the Web. In presentation systems, it is more fascinating when we conduct the presentation stream by the explanation or conversation of presenters. And in order to provide the presentation on demand on the Web, we need character based presentation Agent. There are several ways to control such agents but some are too complex to be used by general users. We try to propose an easy way for the uses. 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 its specification, visual authoring tool, and application of MPML when used as a tool for composing multimodal presentations on the WWW.

Keywords
XML, Markup Language, Multimodal, Presentation Agent

Introduction

Within the past few years, the WWW has been emerging incredibly and becoming an important infrastructure of information service and circulation. One of striking features of the WWW is that people can publicize information to the world easily. Though it is heterogeneous, it is forming a huge space of information. Due to the diversified, complicated and largely scaled contents, the information overload situation is also emerging. A lot of efforts have been made among the researches to overcome this situation and provide better efficiency in information access, fruitful selection and effective usage of information. In many cases, the quality of information depends largely on the content and expression. Unattractive information is likely to be sifted out. Providing attractive and effective information to all ranges of audiences becomes an important matter to information providers.

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. We have provided a solution to this difficulty using visual authoring tool on the upper layer of MPML.

Presentation Agents

Anthropomorphic Agent System

Nowadays, GUIs play an important role in user interface systems. The users are familiar to interaction in form of "See and Point" fashion. According to the achievement of multimedia features, "Ask and Tell" fashion, which enables multimodal interaction and complex input to computers, is currently researched. Among these works, many of them aim to realize communication in "Face-to-Face" fashion, which is the most familiar communication in daily life, by means of anthropomorphous agent interface. The interface is equipped with expressions and movements of computer graphic (CG) characters. Some examples of such multimodal anthropomorphous agent interface are the TOSBURG II by Toshiba [11], the system at Sony CSL [8], VSA system which is embedded onto Web browser [3]-[5] and the system at Electrotechnical Laboratory [6].

Meanwhile, the development is not only the interface itself but also the systems and tools for creating and distributing attractive multimodal information content which will be more and more important from now on. One thing we can expect the most to obtain from the presentation character agent based multimodal information contents is the multimodal presentation. At the moment, there are many research works such as Virtual Human Presenter at University of Pennsylvania, and WebPersona, which have the WWW capability at The German Research Center for Artificial Intelligence (DFKI).

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 Independent**: The content builders usually need to take audiences' OS, browsers and resources into account when providing presentation on 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 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.
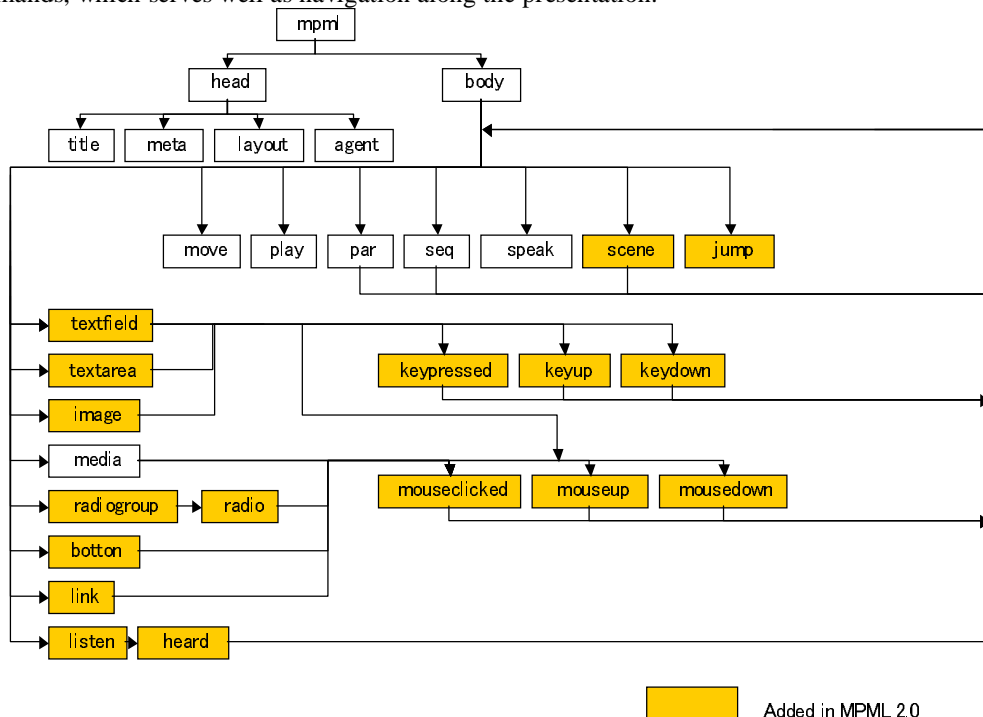

Figure 1: MPML structure tree

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 arbitrary times. The '#PCDATA' in the tree diagram represents text data.

The root of all elements is the tag <mpml>, which has attribute 'id'. The attribute 'id' is utilized to facilitate identification of tags. Most of the tags can be assigned IDs. Moreover the tag <mpml> can include tag <head> or <body> in MPML document. Each tag will be discussed in the next section.

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 by default.

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 the 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 Comparison with Other Markup Languages

The comparison of MPML with other markup languages (SMIL and HTML) is shown in Table 1.

Table 1: 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
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 30	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.
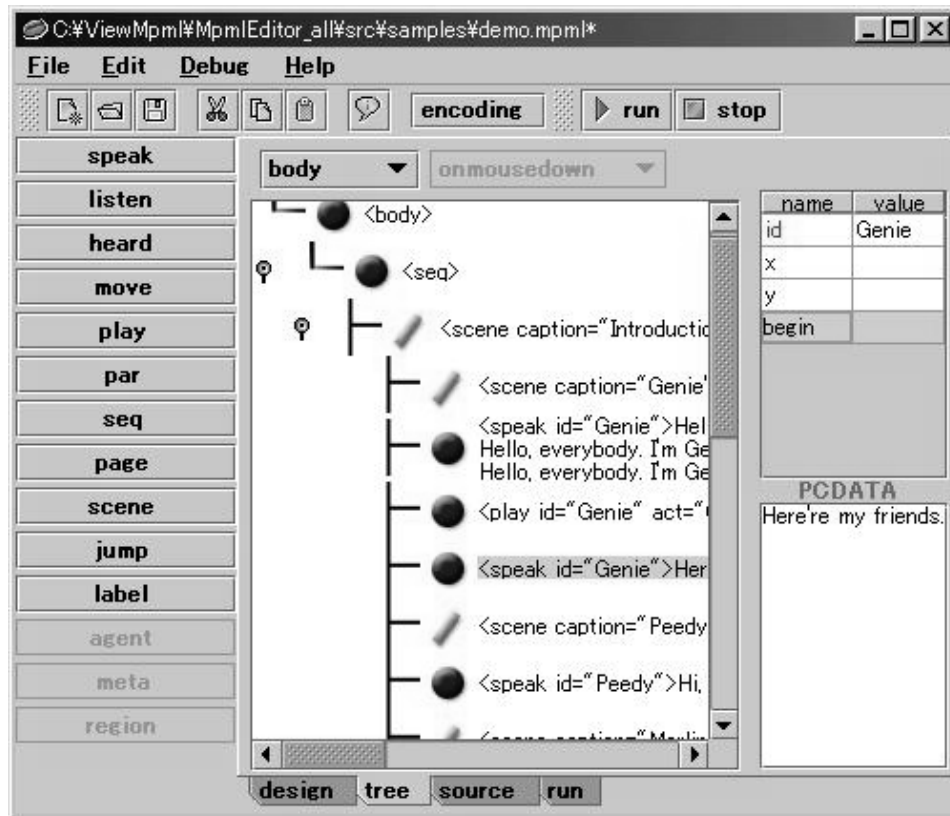
Visual Authoring Tool for MPML

Figure 2: Visual Authoring Tool for MPML.

We have provided a visual authoring tool for MPML as shown in Figure 2. On the left pane, user can choose MPML tags to control the agents and the flow of presentation. The selected tag will be shown in the tree. The user can edit properties of the tag on the right pane. The script can be run by clicking on "run" button. Visual Authoring tool will automatically create MPML script and run from the beginning. Figure 3. shows the look when running composed script. The user can stop the playback at any point of running by clicking on "stop" button. With these facilities, the user can easily create MPML content and publish it on the web in short time.

Concluding Remarks

This paper proposes a visual authoring tool that facilitates the making and distributing of presentation contents with multi-character presentation agent. 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 WWW simply by scripting with the small set of MPML tags. At the moment, some interactive functions, which are sufficient to the presentation aspect, are available.
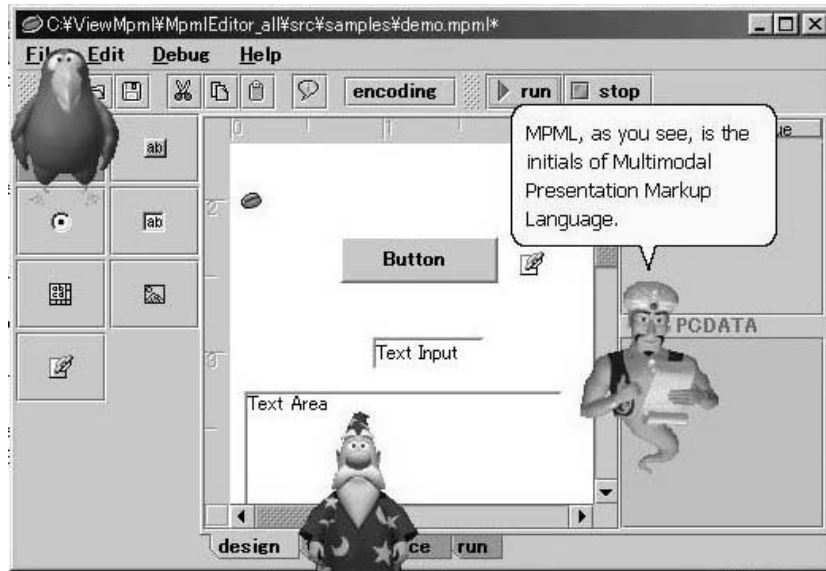
Figure 3: Running MPML Script on Visual Authoring Tool.

REFERENCES

1. 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.

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

3. Dohi, H. and Ishizuka, M.: A Visual Software Agent: An Internet-Based Interface Agent with Rocking Realistic Face and Speech Dialog Function, *AAAI-96 technical report of Internet-Based Information Systems*, WS-96-06, pp. 35-40, 1996.

4. Dohi, H. and Ishizuka, M.: Integration of WWW/Mosaic and Anthropomorphic Interface Agent, *Journal of Electronics, Information and Communication Engineers*, Vol. J79-D-II, No. 4, pp. 585-591, 1996.

5. Dohi, H. and Ishizuka, M.: The construction of Face-to-Face Anthropomorphic Interface Agent, *Journal of Information Engineering*, Vol. 40, No. 2, pp. 547-555, 1999.

6. 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.

7. MPML Home Page:
   http://www.miv.t.u-tokyo.ac.jp/MPML/mpml.html

8. 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.

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

10. SMIL Home Page: http://www.w3.org/AudioVideo/

11. 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.

12. Tsutsui, T., Dohi, H., Ishizuka, M.: The Multimodal Presentation Markup Language MPML for Web Content Presentation by Agent Character, *Technical Report on Virtual Environment Research, MVE98-90*, The Institute of Electronics, Information and Communication Engineers, 1999.

13. Tsutsui, T., Lee, K., Dohi, H., and Ishizuka, M.: The Evolution of Multimodal Presentation Agent, *The 57ᵗʰ Annual Conf. of Information Processing Society of Japan*, No. 1N-9, 1998.

14. XML Home Page: http://www.w3.org/XML/