A Scripting Language for Multi-Character Presentation Agent based on Multimodal Presentation Markup Language

He Binda, 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: This paper proposes an alternative scripting language to control multi-character agent for presentation system. 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 the specification, related tools, and application of MPML when used as a tool for composing multimodal presentations on the WWW.

Introduction

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.

Multi-Character Presentation Agents

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 WWW so that everyone can access the presentation anytime.
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. The audiences cannot feedback their feeling to the content makers.

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.

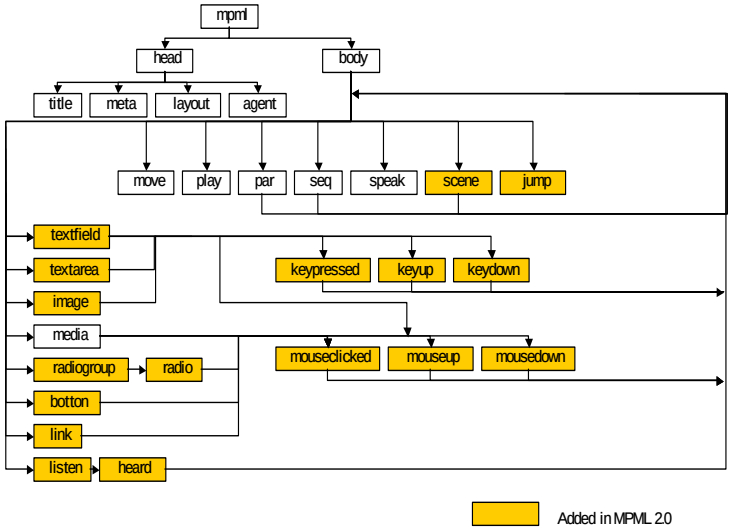

Figure 1: MPML structure tree

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
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.

Concluding Remarks

This paper proposes an alternative scripting language 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.