

Multimodal Presentation Markup Language MPML With Emotion Expression Functions Attached

Yuan ZONG * Hiroshi DOHI Mitsuru ISHIZUKA
Department of Information and Communication Engineering
School of Engineering, University of Tokyo
E-mail: { yzong, dohi, ishizuka } @miv.t.u-tokyo.ac.jp

Abstract

With the increase of multimedia contents in the WWW, multimodal presentation using interactive life-like agents is attractive and becoming important. However, it is not easy for many people to write such multimodal presentations, due to the complexity of describing various behaviors of character agent and their interaction of particular character system with individual (often low-level) description language. In order to overcome this complexity and to allow more people to write attractive multimodal presentation easily, MPML (Multimodal Presentation Markup Language) Version 1.0 has been developed to provide a medium-level description language commonly applicable to many character systems. Also, a new emotion expression function is attached to MPML recently. With this function, we are able to express emotion-rich behavior of the character agent in MPML Version 2.0e. In this paper, We make a simple introduction of MPML Version 1.0, and present the new functions in Version 2.0e. Some multimodal presentation contents are produced in the new version of MPML to show the effectiveness of the new emotion function.

Keyword : *Multimodal presentation, life-like agent, emotions expression*

1. Introduction

The computing industry of the 1990s is in the process of fully adopting the graphical user interface (GUI) metaphor pioneered by Xerox

* Presently with IBM Japan Systems Engineering Co., Ltd.

PARC in the 1970s. This metaphor, first explored by the Smalltalk system on the Alto, was already firmly defined in most significant respects when the Xerox Star was introduced in 1980. Before that, character user interface (CUI) was the main metaphor, which prompts user to type command at the prompt line. The concepts of WYSIWYG editing, overlapping screen windows, and the direct manipulation of system objects as icons had all been thoroughly demonstrated. The subsequent decade has been considerable refinement of the original ideas, particularly regarding usability issues and the idea of visual affordances, but the essence of the original metaphor is intact. As GUIs become the industry standard, it is appropriate to look ahead to the next major metaphor shift in computing. While there are undoubtedly many further improvements that can (and will) be made to GUI metaphor, it seems unlikely that computing in 2010 will still be primarily a process of clicking and dragging buttons and icons on the metaphorical desktop. Improvements in display technology, miniaturization, wireless communication, and of course processor performance and memory capacity will all contribute to the rapid proliferation of the increasingly sophisticated personal computing devices. But it is the evolution of the software capability that will trigger a basic change in the user interface metaphor: computers will become assistants rather than just tools.

The new decade will see increasing efforts to develop software, which can perform large tasks autonomously, hiding as many of the details from the user as possible. The ideal user interface would enable us to perform our tasks without being aware of the interface as the intermediary.

Among many possible Post-GUI interfaces, multimodal user interface is supposed to be the most prominent one. Multimodal user interface uses the character agent as the middle layer between user

and computer, interacting with user and controlling the device and resource of the computer. With the help of the character agents' development, the implement of multimodal user interface becomes possible and effective. Character agent recognizes user's commands through different channels (sound, typing and so forth), and performs the task as user requests. After the task is completed, character agent reports the result by words or actions. With the favor of character agent, user can get information through many information channels (ex. speeches, intonation, emotion, actions and so forth).

One important implementation of utilizing multimodal user interface is multimodal presentation.

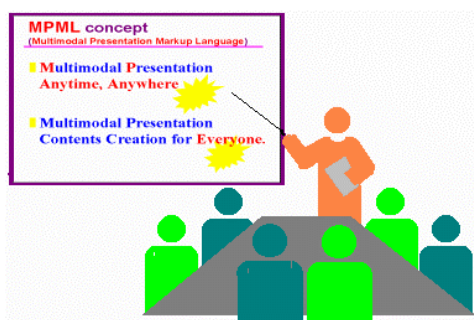


Fig. 1 Current Presentation

Presentation is the process of showing others achievement or approach. With the development of the multimedia technology, presentation technology also developed. Centuries ago, people used text to make appeal. Because text only conveys information through one single modal channel, it was not an effective presentation method. Recently people use presentation tools to make presentation. (ex. OHP, PowerPoint and so forth) As shown in Fig. 1, with the help of these tools, images, movies, text, presenter's sound, intonation convey different information to the audience through different channels respectively. Audience integrates all information from different channels, then understands what happens and what the presenter wants to say completely. Because this method conveys much information through different channels, it is effective and became the most popular presentation method at present.

However, the popularity of this kind of presentation method should not mislead us into thinking that it is an ideal presentation form, just as

the longevity and ubiquity of the now two-decades old graphical user interface should not mislead us into thinking that it is an ideal interface. This presentation method's disadvantage is that the presenter has to be at the meeting hall, which means time and place is restricted. One new presentation method is necessary, which permits presentation process going on anytime and anywhere.

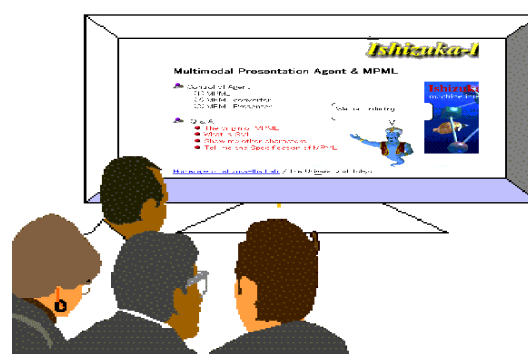


Fig. 2 Presentation Using Life-like agent

Presentation technology develops while interface technology develops. Multimodal presentation appears as the new presentation method. It makes presentation without the restriction of time and place. Fig. 2 illustrates this kind of presentation. The character agents make the presentation instead of human. You can download the presentation contents from WWW, then ask the character agent to make the presentation according the contents.

However, this attractive presentation method didn't replace the current popular PowerPoint presentation tools yet. The reason is that it is too difficult to write the multimodal presentation contents. At present, many character agents exist, and different script language specifications are defined to control different character agents respectively. Most of these script languages are one of a kind (often low-level). It is complex to describe various behaviors of character agent.

In order to overcome the complexity of describing various behaviors of character agents, and write attractive presentation contents easily, we developed MPML (Multimodal Presentation Markup Language).

2. MPML Version 1.0

The goal of MPML (Multimodal Presentation Markup Language) is to make everyone to write attractive multimodal presentation easily.[7]

Most current multimodal presentation contents are written for the particular character system. In most cases, we have to write the detailed description to control the particular agent system. [6]

We hope people can write multimodal presentation easily, just as people can build homepage easily using HTML. So MPML is designed to create multimodal presentation contents independent of the character agents. Some features for MPML Version 1.0 are showed as below:

- *Independent of the character agent system*
MPML is designed to be played by various types of character agent systems, so contents providers don't have to worry about the implement platform when writing MPML script.
- *Easy to describe*
MPML conforms to XML (eXtensible Markup Language), so that everyone, who knows how to write homepage using HTML, should be able to master MPML in a short time.
- *Media synchronization supported*
MPML conforms to SMIL (Synchronized Multimedia Integration Language), which is designed for the synchronization between different types of media. So people can control synchronization between different types of media easily using MPML.
- *Easy control of the character*
With the simple tag description in MPML, contents providers can control character agents' actions easily.
- *Interactive presentation supported*
MPML supports the hyperlink functions. With the combination of the hyperlink function, sound recognition function and sound synthesis function, interactive presentation between audience and character agent is possible.

3. Emotion and MPML Version 2.0e

As the interface layer between the computer and

user, character agent should have not only the communication functions, but also the personality, which will affect the users emotionally. Audience will feel bored after communicating with the agent for a while, if the character agent with a face and a body, however, can only perform machine-like actions. [3,4]

Considering how to attach the personality and sociality to character agent, we focus on the emotion expression functions.[2,5]

Emotion can be expressed as *joy, sadness, anger, surprise, hate, fear* and so forth. The classification category is not defined clearly yet. So we focus on the related research about emotion in cognitive psychology domain.

Cognitive psychology was established in the 1970s. In 1981, Donald Norman identified the topic of emotion as one of twelve major challenges in cognitive science. It was at about this time that three researchers decided to collaborate to explore the extent to which cognitive psychology could provide a viable foundation for the analysis of emotions. Their names are Andrew Ortony, Gerald Clore, Allan Collins. In 1988, they published the book called *The Cognitive Structure of Emotion*, in which they explain the emotion system provided. We took the first letter of their names, and dubbed their emotion structure model as OCC model. [1]

According to OCC's theory, all emotions can be divided into terms according to the emotion-eliciting situations. Emotion-eliciting situations can be divided roughly into 3 types. The first type of emotion-eliciting situation is *consequences of events*. It can be further divided into 2 branches: *consequence for others* and *consequence for self*. The second type of emotion-eliciting situation is *actions of agents*. It is also further divided into 2 branches: *prospects relevant* and *prospects irrelevant*. The third type of emotion-eliciting situation is *aspects of objects*. According to the classification of emotion-eliciting situations, all emotions can be divided into 3 classes, 6 groups and 22 types. Fig. 3 displays the structure of emotions, classified by emotion-eliciting situation, according to OCC's emotion theory.

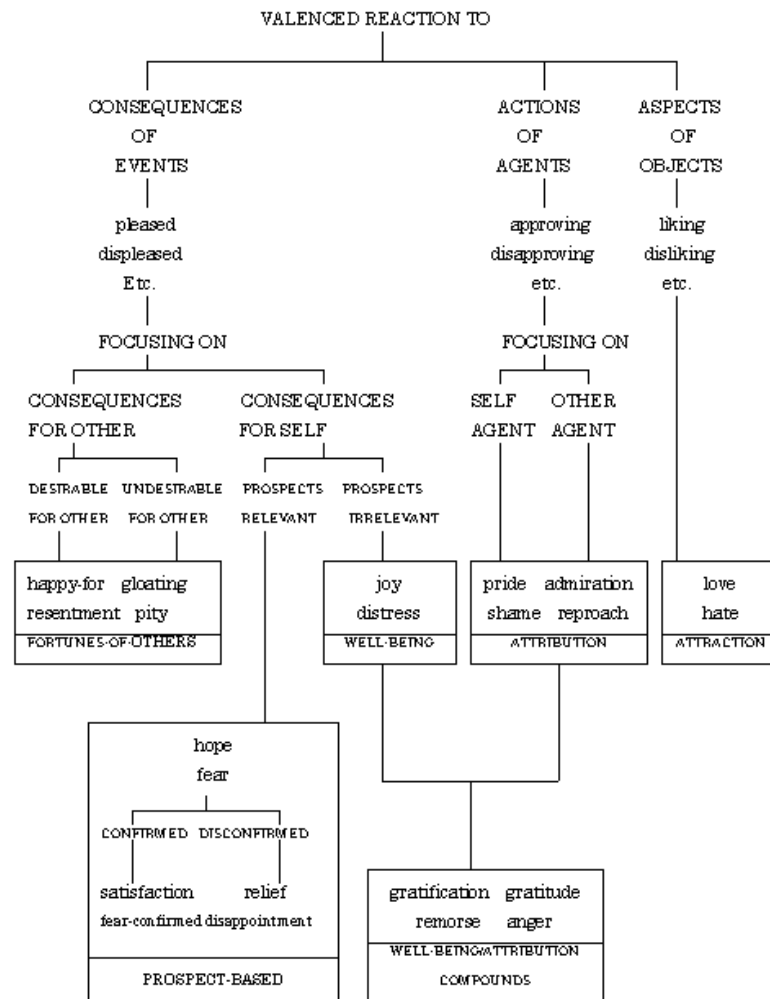


Fig. 3 The Structure of the Emotions

In MPML Version 2.0e, we provide the emotion expression function to control agents' emotion more conveniently. Contents provider can specify 22 types of emotion defined in OCC emotion model, to modify the action performed by character agent. Character agent express the emotion with performing different action, changing the pitch, volume, speed, emphasis of the voice. For example, when the emotion type is specified as "pride", the character agent would wave his hands, then speak loudly with the emphasis at the beginning of the sentence.

Except for the emotion expression function, some new functions are added in Version 2.0e:

- *Page*
Every presentation is divided into pages. The

concept of "page" is equal to the slide in PowerPoint, or OHP presentation. Contents provider needs to describe Multimodal Presentation contents page by page. With the favor of this concept, MPML script became clear and easy to be understood.

- *Fast-forward*
Audience can request jumping from current page to next page or to previous page when watching the presentation. Just as we listen to music, sometime we would like to listen to the same track once again or jump to the next track.
- *Presentation-macro*
Some templates are prepared for particular presentation purpose.

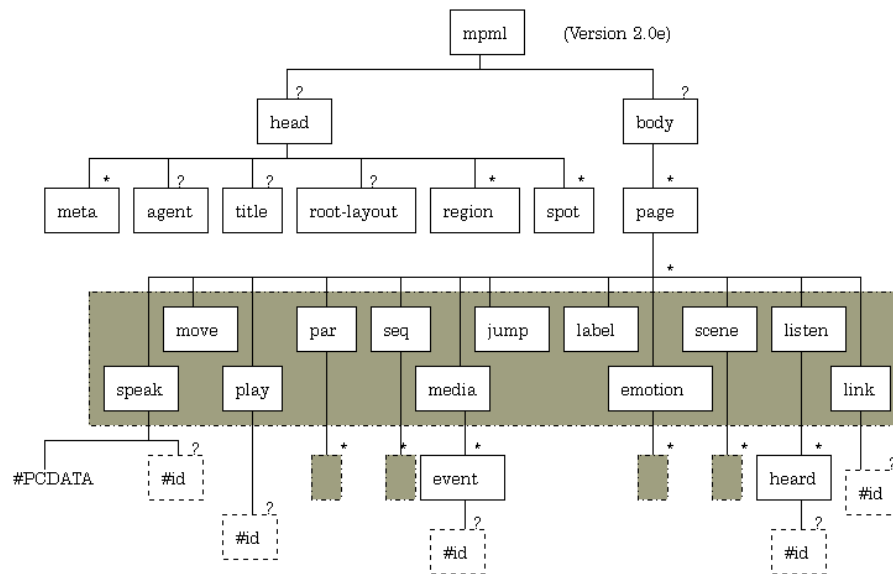


Fig. 4 Tag Structure for MPML Version 2.0e

In order to support the new function in MPML Version 2.0e, some tags are added. Fig. 4 illustrates the tag structure for MPML Version 2.0e. The “?” mark in figure means “more than 1”. The tag type with “?” mark must appear in one MPML script more than once. The “*” mark in figure means “more than 0”. The tag type with “*” mark may not appear in one MPML script. There are 23 types of tag defined in MPML version 2.0e. Comparing with over 80 types of tag in HTML, 23 types of tag is not a large number and should be easy to master.

Below is a sample of MPML script, which use emotion expression function to control character agent.

```
<mpml>
  <head>
    <title> MPML sample </title>
    <meta name="author" content="Zong Yuan" />
    <spot id="spot1" location="600,380" />
    <agent id="PD" character="peedy" />
  </head>
  <body>
    <!-- This is the first page -->
    <page id="first_page" ref="self_intro.html"
      area="right_middle">
      <emotion type="pride">
        <speak>
          My name is Zong Yuan.
```

```
        I am from Tokyo University
      </speak>
    </emotion>
  </page>
  <!-- This is the second page -->
  <page id="second_page" ref="index.html"
    area="right_middle">
    <emotion type="happy-for">
      <speak>
        I am so glad to be able to introduce
        MPML here.
      </speak>
      <move spot="spot1" />
    </emotion>
  </page>
</body>
</mpml>
```

According to the above script, the character agent called "peedy" would give a self-introduction with the "pride" emotion activated at first. Then the agent would state the goal of the presentation – introducing MPML – with the “happy-for” emotion activated.

4. MPML concerning tools

In order to be accepted by most people, authoring tools and audience tools should be provided for

MPML.

As displayed in Fig. 5, two kinds of tools should be provided. One is authoring tools, which help contents provider writing contents. The other is audience tools, which is used to play the multimodal presentation contents.

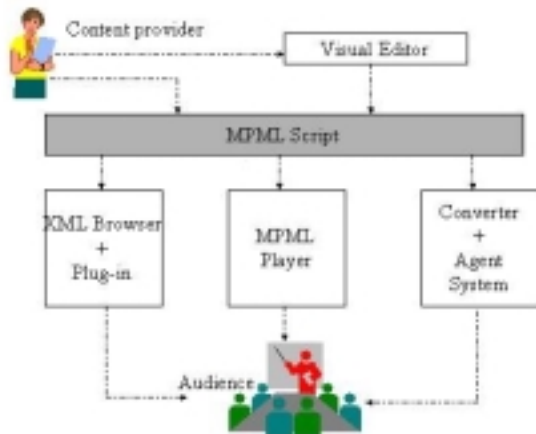


Fig. 5 Tools for MPML

As for the authoring tools, 2 types are considered. One is plain text editor, MPML is easy to learn and write, so that it should be easy to write with the plain text editor. The other authoring tools is visual editor. Just as people use Homepage Builder to built homepage, with the help of visual editor for MPML, people can built the multimodal presentation content without the knowledge of MPML. Also, it will help contents provider to write attractive contents more easily and faster. The visual editor for MPML is under construction.

Audience tools is also necessary for audience to watch the multimodal presentation. 3 types of audience tools are considerable and have been developed already. First is MPML player. One player called *ViewMpm* was developed for MPML version 2.0e already. Fig. 6 displays the screen when *ViewMpm* works. It supports all tags defined in MPML Version 2.0e's specification. The second tool type is a converter which converts MPML script to the script what are understood by particular agent system. At present two kinds of converters are already developed for MPML Version 1.0 (an old version of MPML). And the third tools type is XML-browser with plug-in program [8]. Since MPML conforms to XML, it can be understood by XML-browser. Internet Explorer 5.0 is one of

XML-browsers. Now, one plug-in program written by XSL (XML Style Language) for Internet Explorer 5.0 is already developed.

All specifications and executive programs for the newest version of MPML (MPML Version 2.0e) are available on WWW. Everyone can download them from the following site:

- Homepage written in English:
<http://www.miv.t.u-tokyo.ac.jp/MPML/en/2.0e/>
- Homepage written in Japanese:
<http://www.miv.t.u-tokyo.ac.jp/MPML/jp/2.0e/>

The instruction to install MPML player (*ViewMpm*) is published in the above pages. Some attractive contents are also included. A movie file around 1.4 Mbytes for a 15 seconds multimodal presentation is available at the following site:
<http://www.miv.t.u-tokyo.ac.jp/MPML/en/2.0e/movies/mpmlmovie.mpg>

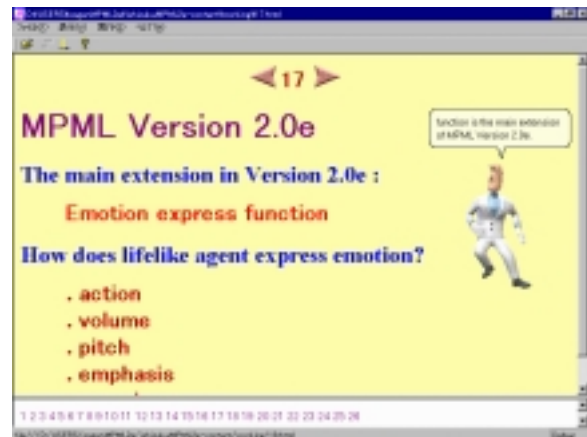


Fig. 6 MPML Player (ViewMpm)

5. Conclusion

The goal of MPML is to let most people create multimodal presentation contents easily.

In MPML Version 2.0e, we keep the features of version 1.0 and applied some new functions to MPML. The main extension is emotion expression function, which conforms to OCC emotion models. For future potential change of the emotion models, this system didn't restrict emotion expression function to OCC models. We can change the

emotion model easily by changing the text setting files, which means enough scalability is given to emotion expression function attached to MPML.

Because the current existing character agents were not designed for emotion expression when they were developed, they can not express emotion fully and exactly as we desired. As for the future work, we are going to use the customized 3D character agent to express emotion more freely and naturally.

On the other hand, OCC emotion model defined some emotion-eliciting situation and some types of emotions elicited by these situations. In MPML Version 2.0e, we only provide the emotion types defined in OCC emotion model to MPML. In the future, the emotion-eliciting situation's input will also be considered.

Reference

- [1] A. Ortony, G. L. Clore and A. Collines: *The cognitive Structure of Emotions*, Cambridge Univ. Press(1988)
- [2] *Proc. Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, IEE Computer Society Press. Los Alamitos, Calif. , 1999
- [3] M. Elson, "The Evolution of Digital Characters," *Computer Graphics World*, Vol. 22, No. 9, Sept. 1999, pp. 23-24
- [4] F. Thomas and O. Johnson, *Disney Animation: The Illusion of Life*, Abbeville Press, New York, 1981
- [5] Nagao, K. and Takeuchi, A. : Speech Dialogue with Facial Displays, *Multimodal Human-Computer Conversation 32nd Annual Conf. Of ASSOC of Computational Linguistics*, 102-109, 1994
- [6] <http://msdn.microsoft.com/workshop/imedia/agent/>
- [7] <http://www.miv.t.u-tokyo.ac.jp/MPML>
- [8] <http://www.w3.org/TR/REC-xml/>