

Emotion Expression Function in Multimodal Presentation

Yuan Zong¹, Hiroshi Dohi², Helmut Prendinger², and Mitsuru Ishizuka²

¹ IBM Japan Systems Engineering Co., Ltd.,
1-1, Nakase, Mihama-ku, Chiba-shi, Chiba 261-8522, Japan
Tel: 81-43-297-6055, Fax: 81-43-297-4836
yzong@jp.ibm.com

² Department of Information and Communication Engineering,
School of Engineering, University of Tokyo,
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
Tel: 81-3-5841-6755, Fax: 81-3-5841-8570
{dohi, helmut, ishizuka}@miv.t.u-tokyo.ac.jp

Abstract. With the increase of multimedia content on the WWW, multimodal presentations using interactive lifelike agents become an attractive style to deliver information. However, for many people it is not easy to write multimodal presentations. This is because of the complexity of describing various behaviors of character agents based on a particular character system with individual (often low-level) description languages. In order to overcome this complexity and to allow many people to write attractive multimodal presentations easily, MPML (Multimodal Presentation Markup Language) has been developed to provide a medium-level description language commonly applicable to many character systems. In this paper, we present a new emotion function attached to MPML. With this function, we are able to express emotion-rich behaviors of character agents in MPML. Some multimodal presentation content is produced in the new version of MPML to show the effectiveness of the new emotion expression function.

1 Introduction

An interface is a necessary part of human-computer interaction. The ideal user interface would let us perform our tasks without being aware of the interface as the intermediary. The longevity and ubiquity of the now two-decade old graphical user interface should not mislead us into thinking that it is an ideal interface.

Among many possible Post-GUI interfaces, a multimodal interface is supposed to be the most promising one. A multimodal interface uses the character agent as the middle layer between user and computer, interacting with user and controlling the

device. The character agent recognizes the user's command and runs a task as the user requests. After the task is completed, the character reports the result by using verbal output or actions. By employing an character agent, the user can get information from many information channels (e.g., speech with intonation, emotion, actions and so forth).

One important implementation of a multimodal interface is multimodal presentation, which is an attractive way to present research work or products. With the development of the multimedia technology, presentation technology evolved. Centuries ago, people used text to appeal the audience. Because text only conveys information through a single channel, it was not a very effective presentation method. Recently, people use various presentation tools to make presentation (e.g., OHP, PowerPoint, and so forth).



Fig. 1. Current Presentation Agent

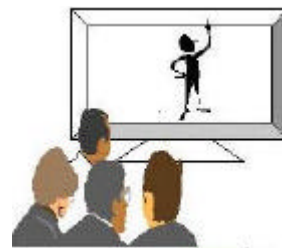


Fig. 2. Presentation with Lifelike Agent

As shown in Fig.1, multimodal presentation allows conveying different information through different channels, such as images, movies, text, and presenters' speech. Because this presentation method conveys information through different channels, it is more effective and became the most popular presentation method at present. However, its disadvantage is that the presenter has to be at the meeting hall, thereby restricting the presentation to a certain time and place.

The solution is Multimodal presentation. It is a new presentation method to make presentation without the restriction of time and place. Fig. 2 illustrates this kind of presentation. The character agents make the presentation instead of a human presenter. You can download the presentation content from the WWW, then ask character agent make the presentation according the content.

However, this attractive presentation method did not replace the current popular PowerPoint presentation tools yet. The reason is that it is too difficult to write the multimodal presentation content. There are many character agents, and different script language specifications are defined to control different character agents, respectively. Most of these script languages require rather low-level programming skills.

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

2 MPML 1.0

The goal of MPML (Multimodal Presentation Markup Language) is to enable everyone to write attractive multimodal presentation easily [9]. Current multimodal presentation content is mostly written for a particular character system. In many cases, one has to program a detailed description to control the particular agent system [8].

We envision that people can write multimodal presentations easily, just as people can write homepage easily using HTML. So MPML is designed to write multimodal presentation content independent of specific character agents.

Some features of MPML Version 1.0 are:

- *Independent of the Character agent system.*
- *Easy to describe*, i.e., anyone who understands HTML should be able to learn MPML in short time.
- *Media synchronization supported*, because MPML conforms to SMIL.
- *Easy to control character*
- *Interactive presentation supported*

3 Emotion and MPML 2.0e

As the interface layer between computer and user, a character agent should not only have communication abilities, but also personality traits which let users feel affection. If a character agent, which has the face and body, but can only perform machine-like reactions, the audience will soon feel bored when communicating with the character agent [7,2]. Considering personality and social behavior of the character agent, we focus on emotion expression functions [4,6].

Emotion can be expressed as *joy, sadness, anger, surprise, hate, fear* and so forth. There is no generally accepted classification system of emotions yet. So we focus on research about emotions in cognitive psychology. In 1988, Andrew Ortony, Gerald Clore, and Allan Collins published a book called *The Cognitive Structure of Emotion*, in which they provide a detailed analysis of emotions [5]. Their analysis became well known as the OCC model.

According to the OCC model, all emotions can be divided into terms according to the emotion-eliciting situation. Emotion-eliciting situations can be divided roughly into 3 types. The first type of emotion-eliciting situation is *consequences of events*. The second type of emotion-eliciting situation is *actions of agents*. The third type of emotion-eliciting situation is *aspect of objects*. According to the classification of emotion-eliciting situations, all emotions can be divided into three classes, six groups and twenty-two types of emotion (Fig.3).

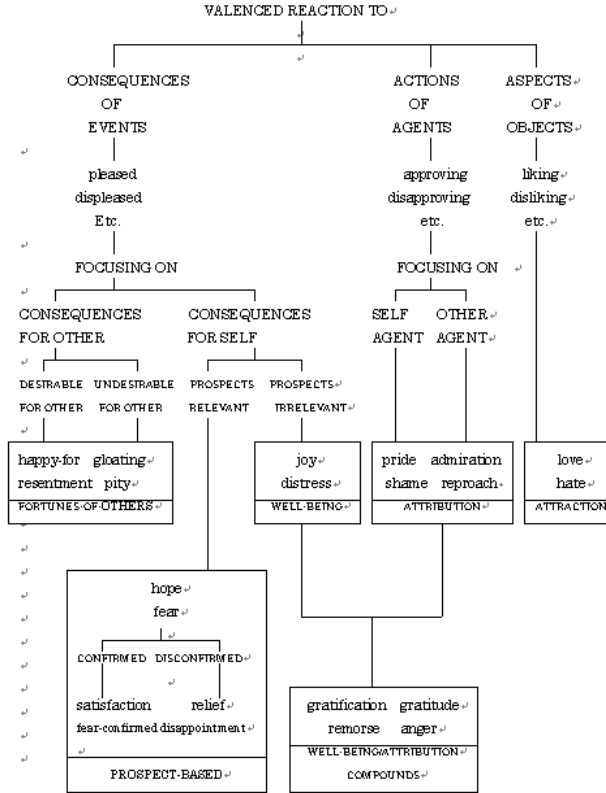


Fig. 3. The Cognitive Structure of Emotion

In MPML Version 2.0e, we provide an emotion expression function to control agents' emotion more conveniently. The content provider can specify the twenty-two types of emotion defined in OCC emotion model, and accordingly modify the action performed by character agent. The character agent expresses the emotion by performing different actions and changing speech parameters (pitch, volume, speed, and emphasis of certain words). 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 functions, some new functions are added in Version 2.0e:

- *Page*: Every presentation is divided into individual pages. Content providers may describe content page by page.
- *Fast-forward*: The audience can request to go to the next or previous page when watching the presentation.
- *Presentation-macro*: Some templates are prepared for particular presentation purposes.

Fig. 4 illustrates the tag structure for MPML Version 2.0e.

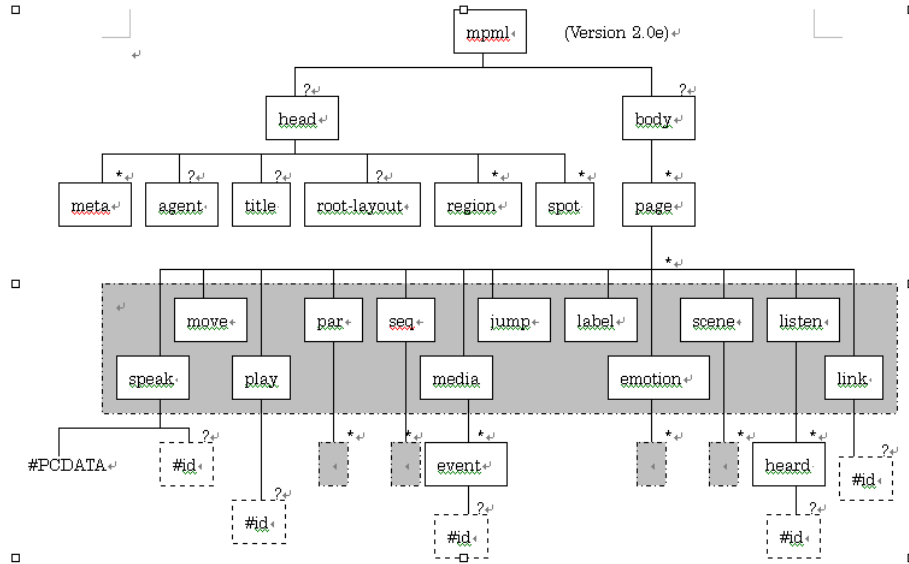


Fig. 4. Tag Structure of MPML

The below is a sample for MPML script:

```

<mpml>
<head>
  <title> MPML Presentation </title>
  <agent id="PD" character="peedy">
</head>
<body>
  <page id="first" ref="self_intro.html">
    <emotion type="pride">
      <speak>
        My name is Zong Yuan,
        I am from Tokyo University.
      </speak>
    </emotion>
  </page>
</body>
</mpml>

```

According to the above script, the character agent called “peedy” would give a self-introduction with the “pride” emotion activated.

4 Tools for MPML

In order to be accepted by many people, authoring tools and audience tools should be provided for MPML. As for authoring tools, two types are conceivable. One is a plain text editor. Since MPML is easy to learn and write, it should be easy to be written with a plain text editor. Another authoring tool might be a visual editor. Just as people use Homepage Builder to built homepages - with the help of a visual editor for MPML - people can script multimodal presentation content without the knowledge of MPML. A visual editor for MPML is under construction.

Audience tools are also necessary for users to watch the multimodal presentation. Three types of audience tools are considered and have been developed already. One type is the MPML player. One player called “ViewMpml” was developed for MPML 2.0e already. The second tool type is a converter which converts MPML to a script that is understood by a particular agent system. At present two kinds of converters are already developed for MPML 1.0 (an older version of MPML). The third tool type is an XML browser with plug-in [10]. Because MPML conforms to XML, it can be understood by an XML browser. At present, one plug-in program written in XSL has already been developed.

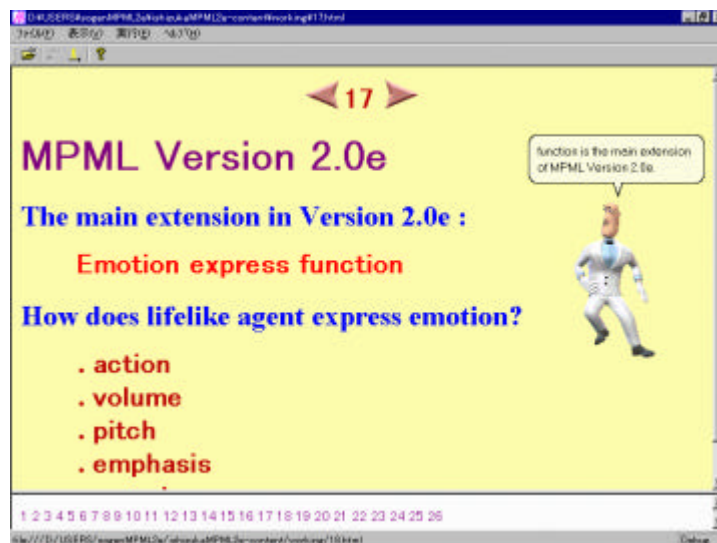


Fig. 5. MPML Player (ViewMpml)

Fig. 5 displays ViewMpml, a MPML player developed for MPML Version 2.0e. It supports all tags defined in MPML Version2.0e's specification. It is free and can be download from the following site:

<http://www.miv.t.u-tokyo.ac.jp/MPML/en/2.0e/>

Moreover, a movie file (1.4 Mbytes) for a 15 seconds multimodal presentation is provided at the next site:

<http://www.miv.t.u-tokyo.ac.jp/MPML/en/2.0e/movies/mpmlmovies.mpg>

5 Conclusions

The goal of MPML is to enable many people to publish multimodal presentation content easily. In MPML Version 2.0e, we keep the features of Version 1.0 and applied some new functions to MPML. The effectiveness of using character agents for presentation relies on the so-called “persona effect”, which says that the mere presence of an animated character makes presentations more enjoyable and effective [3]. One of our main goals was that presentations can be run anytime and anywhere. In particular, presentations should be run client-side in a web-browser (Microsoft Internet Explorer, 5.0 or higher). This restriction ruled out other possibilities, such as running pre-recorded video-clips, since they have long loading times and are expensive to produce. However, we are aware that experiments suggest video-recordings of real people to be the most effective presentation method (except for human presentation performance, of course).

The main improvement of Version 2.0e is an emotion expression function, which integrates the emotions identified in the OCC model to MPML. The mapping from the emotions to the character agent’s behavior (action and speech) is done by common sense (intuition) rather than according to empirical investigation. However, we can change the emotion parameters easily by changing the text setting files. A prime candidate would be the work on “basic emotions” [1], which identifies a set of emotions that have distinctive signals (e.g., distinctive facial expressions or distinctive speech).

The currently available character agents were not designed for emotion expression. Therefore we started developing customized 3D character agents to express emotion more freely and naturally. Another idea is to let the character agent reason about the emotion-eliciting situation.

References

1. Ekman, P.: An Argument for Basic Emotions. *Cognition and Emotion*, 6, 3-4, 1992, 169-200
2. Elson, M.: The Evolution of Digital Characters. *Computer Graphics World*, Vol. 22, No. 9, Sept. 1999, 23-24
3. Lester, J.C., Converse, S.A., Stone, B.A., Kahler, S.E.: *Animated Pedagogical Agents and Problem-solving Effectiveness: A Large-scale Empirical Evaluation*. Artificial Intelligence in Education, IOS Press: Amsterdam, 1999, 23-30

4. Nagao, K., Takeuchi, A.: Speech Dialogue with Facial Displays. Multimodal Human-Computer Conversation. 32nd Annual Conference of the Association of Computational Linguistics. 1994, 102-109.
5. Ortony, A., Clore, G.L., Collins, A.: The Cognitive Structure of Emotions. Cambridge Univ. Press, 1988.
6. Proceedings Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems. IEEE Computer Society Press. Los Alamitos, CA, 1999.
7. Thomas, F., Johnson, O.: Disney Animation: The Illusion of Life. Abbeville Press, New York, 1981.
8. <http://msdn.microsoft.com/workshop/imedia/agent/>
9. <http://www.miv.t.u-tokyo.ac.jp/MPML>
10. <http://www.w3.org/TR/REC-xml/>