

# MPML 2.0e : Multimodal Presentation Markup Language supporting Emotion Expression

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 rapid progress of 3D animated character agent technology, multimodal presentation using interactive lifelike agents is attractive and is becoming more and more important. Various types of lifelike agents have already been developed and are playing important roles in effective multimodal presentations. However, it is not easy for many people to write multimodal presentations, because of the complexity of describing various behaviors of character agents based on a particular character system with individual (often low-level) description language. In order to overcome this complexity and allow more 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. At the same time, we noticed the importance of emotion expression in a prolonged presentation using lifelike agent. So we applied the emotion expression function in the new version of MPML.*

*In this paper, we make a simple introduction of MPML, and then present a new emotion function attached to MPML version 2.0e. With this function, we are able to express emotion-rich behaviors of the character agent in MPML. Some multimodal presentation contents are produced in the new version of MPML to show the effectiveness of the new emotion expression function.*

**Keyword:** *multimodal presentation, lifelike agent, emotion expression*

## 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.

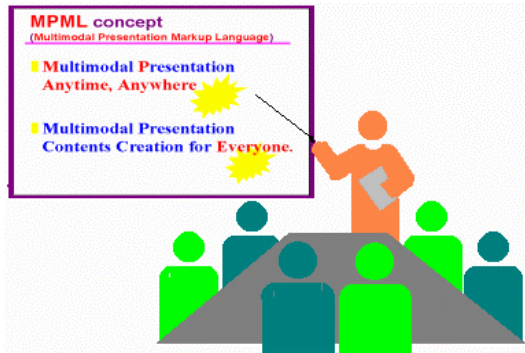
Among many possible Post-GUI interfaces, a multimodal interface is supposed to be the most promising one. A multimodal interface uses the character agent as the middle layer between user and computer, interacting with user and controlling the device. The character agent recognizes the user's command 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 effective and attractive presentation method.

Presentation is a 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 make presentation under the favor of various presentation tools (e.g., OHP, PowerPoint, and so forth).

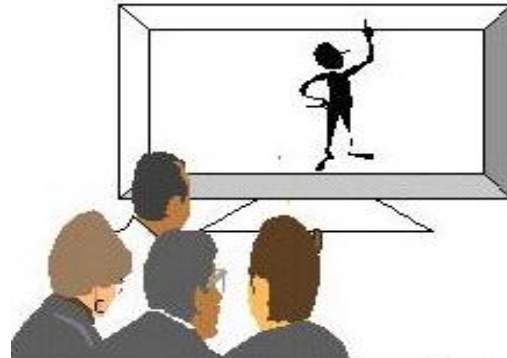
As shown in Fig.1, current presentation, using assistive tools, allows conveying different information through different channels, such as images, movies, text, and presenters' speech. Because this presentation method conveys information through different channels, it is more effective and became the 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.



**Fig. 1 Current Presentation**

To avoid the restriction of time and place, people run some pre-recorded video-clips to make presentation instead. However, video-clips are expensive to produce and don't support the interactive information exchange between presenter and audience. Therefore, new presentation method, as effective as current presentation method using presentation tools by human, without the restriction of time and place, is necessary.

The solution is Multimodal presentation. It is a new method to conduct a 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, and then ask character agent to make the presentation according to the content.



**Fig. 2 Presentation Using Life-like Agent**

However, this attractive presentation method can 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.

One of the most famous examples of multimodal presentation using character agent is the "Virtual Human Presenter" [11] developed by researchers in University of Pennsylvania. It is an auto weather forecast system, using a 3D CG character agent called Jack. Using the command-embedded script, we can control Jack to do weather forecast. Most multimodal presentation agent systems, including the above example, use different script language to control their agents.

In order to overcome the complexity of describing various behaviors of character agents, and to write attractive presentation content easily, we have 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 [8,11].

We envision that people can create

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

Last year, MPML Version 1.0 was developed for above purposes. Some features of MPML Version 1.0 are:

- *Independent of the Character agent system.*  
MPML is designed to let content provider get free from considering distinction between various character agents. Most current character agent controlling languages are for particular character system, so that content provider has to make different script for different character agent system respectively, even for the same presentation. Using MPML, content provider doesn't have to consider which agent system would be used to run presentation, when writing content. MPML realizes "write once, run anywhere", and makes content provider to concentrate on the presentation material itself.
- *Easy to describe*  
MPML conforms to XML. It is a simple language, designed only for multimodal presentation. It provides a set of tags to control the presentation. There are around 20 types of tags, much less than those in HTML. Anyone who understands HTML should be able to master MPML in short time.
- *Media synchronization supported,*  
MPML is in conformity with SMIL (Synchronized Multimedia Integration Language) [12], which is defined for media synchronization. Because of this conformity, Multimodal presentation is effective especially in presentations consisting of various types of media

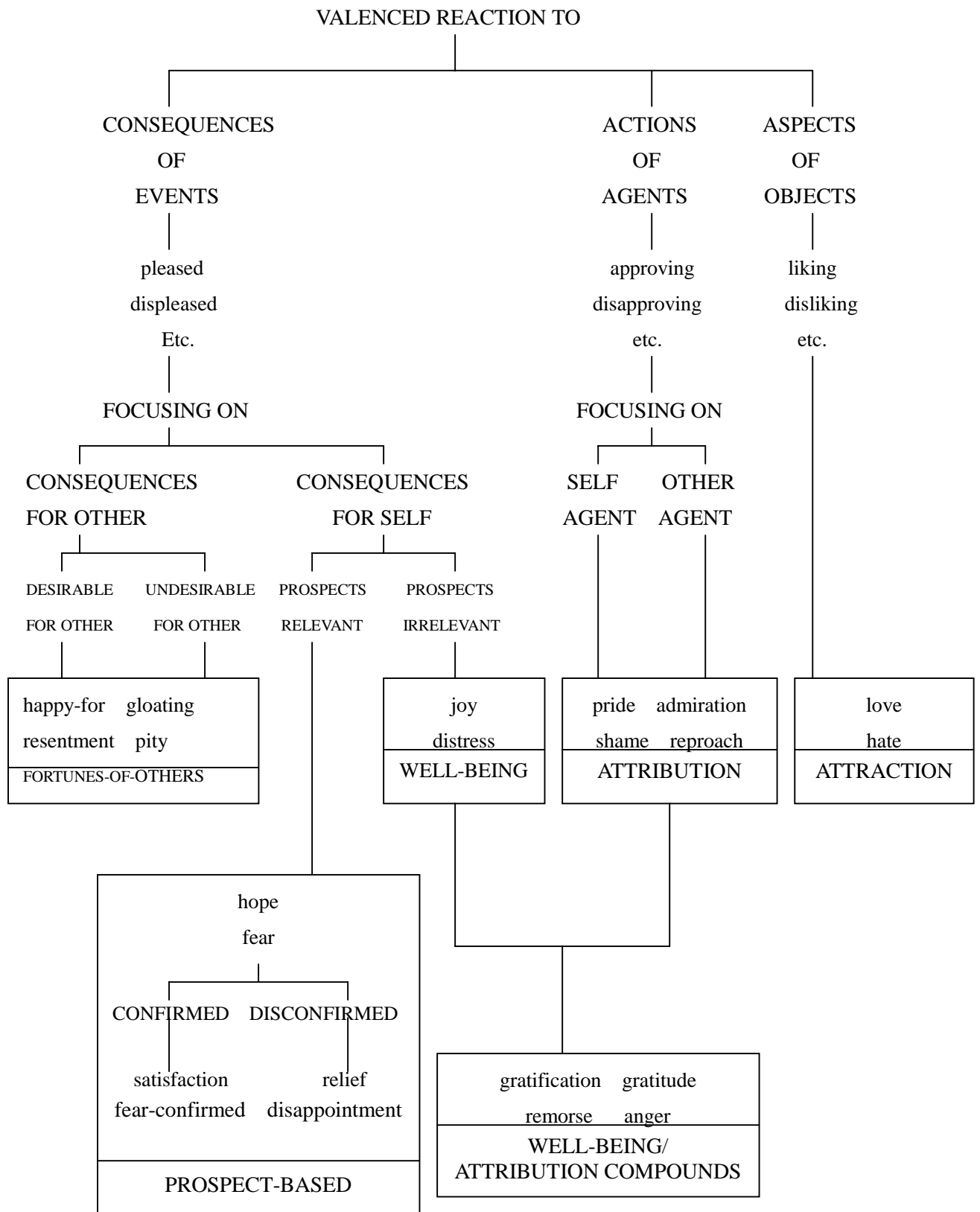
(e.g., text, image, voice and so forth). Synchronization between different types of media is very important. Using MPML, media synchronization is easily realized.

- *Easy control to character*  
Character agent makes presentation instead of human. Multimodal presentation content must control agent to act as human during the presentation process. MPML provides 3 types of tags to control agent's speech, action and location.
- *Interactive presentation supported*  
MPML supports interactive presentations, because information exchange is a key component during presentation. And this overcomes the difficulty pre-recorded video clips face, lack of interaction.

### 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 lets users feel affection. If a character agent, who has the face and body, but can only perform machine-like actions, 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.



*Fig. 3 The Structure of the Emotion*

OCC model is well-known in cognitive psychology domain, and widely cited in persona affection concerning research.

According to the OCC model, emotion is elicited by circumstance conditions and human's current status. Circumstance conditions may be any event taken place. Subject to the same event, the elicited emotion may be different according to human's current status. For example, one student in a class got a scholarship. His classmates will have different elicited emotion. The student's friend will feel happy and pride of the student. However, the other students perhaps will be envious. In OCC model, circumstance conditions are called emotion-eliciting situation. 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).

In order to attach personality and sociality to character agent, we focus on OCC model, and decided to apply emotions, which are defined in OCC model, to MPML.

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 and develop content in a unit of page. This function is especially useful for prolonged multimodal presentation.
- *Fast-forward:*  
We provide the fast-forward and fast-reverse function in MPML Version 2.0e. The audience can request to go to the next or previous page when watching the presentation. Content provider will also profit from this function, when debugging their content.
- *Presentation-macro:*  
Some templates are prepared for particular presentation purposes. With these templates, content provider would develop multimodal presentation content faster and more easily.

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

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.

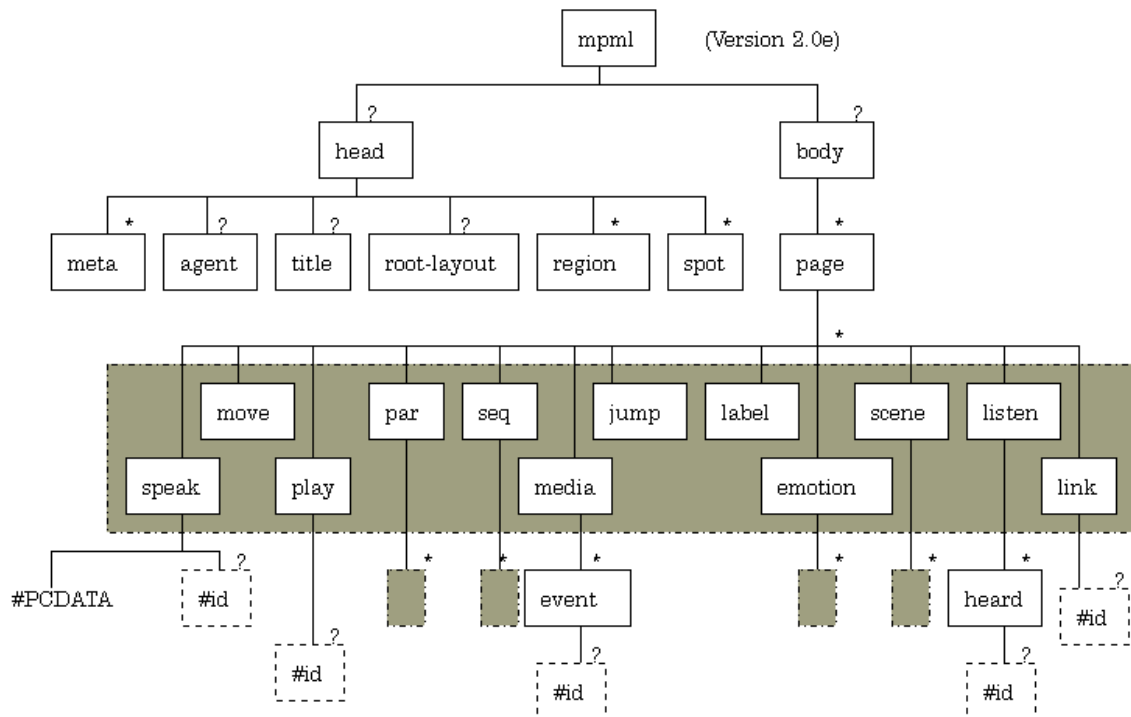


Fig. 4 Tag Structure for MPML 2.0e

#### 4. MPML related tools

In order to be accepted by general users, 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 compose 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 that 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 is in conformity with XML, it can be understood by an XML-browser. At present, one plug-in program written in XSL has already been developed.

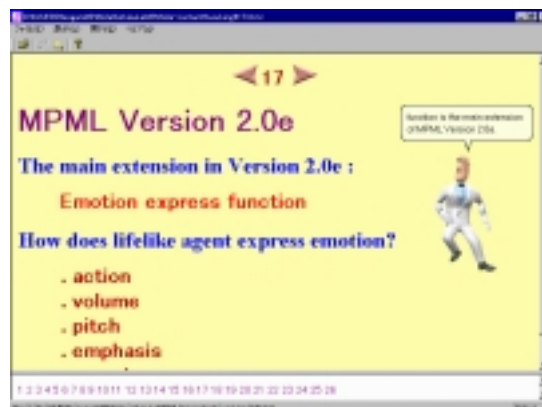


Fig. 5 Mpml Player (ViewMpml)

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 downloaded 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 available at the next site:

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

## 5. Conclusion

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/>
11. Noma, T. and Badler, N.I.: "A Virtual Human Presenter", *Proceeding IJCAI-97 Workshop on animated Interface Agent: Making them Intelligent*, Nagoya, Japan, 1997, pp. 45-51
12. <http://www.w3.org/AudioVideo/>