Humanoid Robot Presentation Controlled by Multimodal Presentation Markup Language MPML

Yasubumi Nozawa[1], Hiroshi Dohi[2], Hitoshi Iba[1], Mitsuru Ishizuka[3]

[1] School of Frontier Sciences, University of Tokyo
[2] School of Engineering, University of Tokyo,
[3] School of Information Science and Technology, University of Tokyo
7-3-1, Hongo, Bunkyo-ku,
Tokyo 113-8656, JAPAN

E-mail   *{nozawa,iba}@iba.k.u-tokyo.ac.jp*
         {dohi,ishizuka}@miv.t.u-tokyo.ac.jp

Abstract

We have developed a Multimodal Presentation Markup Language, called MPML. In our previous studies, we have succeeded to make attractive multimodal presentation with animated virtual characters easily. Then we have combined the MPML with a two-legged humanoid robot, instead of the animated character on 2D screen. It enables an end-user to control freely the humanoid robot presenter for his/her own web-based multimodal presentation. The humanoid robot introduces the multimedia contents with a voice with pointing at a screen using a laser pointer. A single MPML program can generate both animated character presentation on 2D screen and humanoid robot presentation in 3D space. We also show empirically how controllable and expressive the presentation is by means of the humanoid robot.

Figure 1   Humanoid Robot Presentation

1   Introduction

Life-like agent interface is becoming increasingly important as an information advisor, a personal assistant, news presenter, and so on.

Many languages and tools for controlling life-like characters have been developed. The VHML (Virtual Human Markup Language) [4] and the CML (Character Markup Language) [1] are both scripting and representation languages for animation. The APML (Affective Presentation Markup Language) [2] targets communicative functions. The TVML (TV program Making Language) [11] and our MPML (the Multimodal Presentation Markup Language) [10] are the scripting language for presentations.

The MPML is the markup language, and it is designed for the user without programming skills to make his/her own multimodal presentation with the life-like agent easily.
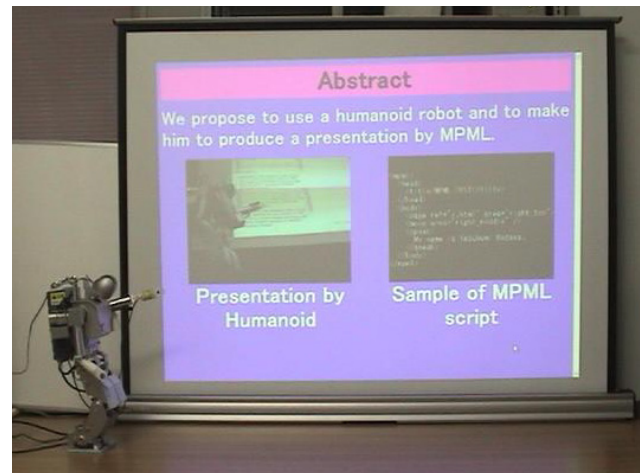
These systems, including our MPML, have used the animated virtual character as the life-like agent apparently living on the screens of computers.

Recently interactive life-like robots are also brought into our domestic environments. It is well known that actions of a physical robot give strong impression to users. When the robot raises his hand, it will atract all users' attention around the robot. Some amusement or entertainment robots have been developed like the AIBO etc, and the opportunities to touch and communicate with the pet robot become increasing. However, in general, it isn't allowed for an end-user to control freely the physical robot except for a few simple cases. Each robot has own different control method, and only the special programer who has internal knowledge can write the control program.

In this paper, we propose a new web-based multimodal presentation system with a humanoid robot presenter. As a presenter, we use the small two-legged

humanoid robot, instead of the animated character. The humanoid robot presentation controlled by the MPML is shown in Figure 1.

2 MULTIMODAL PRESENTATION MARKUP LANGUAGE, MPML

2.1 The MPML features

Humans communicate using not only language, but also gesture, contact action, semi-language, emotion expressionm and so on.

We have developed a multimodal presentation markup language called MPML. The appearance of HTML made an end-user that it could send the personal information freely. The MPML enables an end-user to describe an attractive multi-modal presentation easily with it. In our previous studies, we have succeeded in controlling animated characters by using the MPML.

The MPML is an XML-based scripting language designed for multimodal presentation with the character agent.

The salient features of the MPML are follows,

- **Easy descriptiveness** The MPML scripting using some tags is intuitive and is not assume programming skills. Anyone can describe the script easily with some knowledge that is the same as writing an HTML script.
- **Character control function** The MPML has character control function. It mainly uses the Microsoft Agent package as the animated virtual character. Some versions support other characters, the VRML or a 3D character with MS agent interface.
- **WWW-based presentation** The hyperlink function is supported. The MPML uses the ordinal HTML file as a background scene of the presentation. The HTML file can include movie and sound data. Speech synthesis and speech recognition function is also available.
- **Easy distribution** The MPML script itself is distributed on the WWW.

Therefore, it is easy to write a script of an attractive presentation that the animated character introduces the contents that are written in HTML format using voice and gesture.

2.2 A simple example of the MPML presentation

The MPML programming is easy. Figure 2 is a simple but complete example of the MPML with the animated character. (The MPML families have some versions and dialects. This example is available on the MPML ver2.0e system only) [8].

```
<mpml>
  <head>
    <title>MPML Homepage</title>
    <agent id="PD"
           character="peedy"/>
  </head>

  <body>
    <page ref="top.html">·········①
      <play act="greet" /> ········②
      <move x="400" y="200" />····③
      <play act="GestureRight" />
                          ·······④
      <speak>
        This is MPML Homepage. }··⑤
      </speak>
    </page>
  </body>
</mpml>
```

Figure 2   A simple example of the MPML

Basic tags are summarized as follows.

- **Page tag**

e.g.) **<page ref="*top.html*"> ... </page>**

The **<page>** tag shows the page break of the presentation. The "**ref=**" argument specifies a URL. This URL page is used as a background scene of the presentation. The single MPML file includes some **<page>** tags.

- **Play tag**

e.g.) **<play act="*greet*" />**

The **<play>** tag invokes a specified action. All actions are registered in advance. The "**act=**" argument specifies the action. For examples, the "GestureRight" action makes the character pointing to the right by hand gesture. The name of the action is character-dependent.

- **Move tag**

e.g.) **<move x="*400*" y="*200*" />**

The **<move>** tag moves the animated character to the position on the vertical screen. It is often used with the **<play>** tag. The character jumps to the proper position and then points out an important word or object displayed on the screen.

- **Speak tag**

e.g.) **<speak>** *Hello, World* **</speak>**

The text surrounded by the **<speak>** and **</speak>** tags is synthesized by a Text-to-Speech engine.
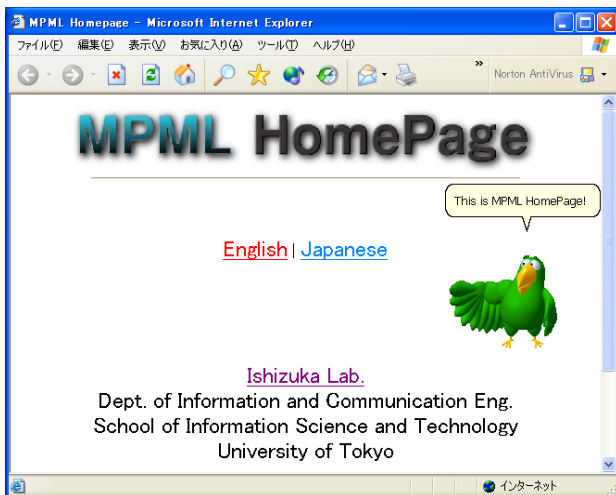
Figure 3 An example of the MPML presentation. The animated character "Peedy" appears and introduces the web page.

The script of Figure 2 gives the following presentations. Figure 3 shows the screen image.

① The "top.html" page is opened
② The animated character "peedy" appears and greets. "Peedy" is a parrot character of the Microsoft Agent.
③ Peedy moves to the position (400, 200) on the screen, and
④ points to the right by hand gesture, and
⑤ says "This is MPML HomePage."

In addition, the MPML has other tags.

The **`<listen>`** tag is supported for a simple speech recognition function.

```
e.g.) <listen>
        <heard word="System">…</heard>
        <heard word="Action">…</heard>
      </listen>
```

Speech recognition function becomes available at this point.

At present, speech recognition doesn't reach the practical use level yet. Then it is mainly used for the selection of the presentation topics.

The **`<emotion>`** tag modulates actions and changes attributes of the synthesized voice.

```
e.g.)   <emotion type="happy-for">
          ...
        </emotion>
```

New actions and voice attributes are added automatically by a specified state of emotion.

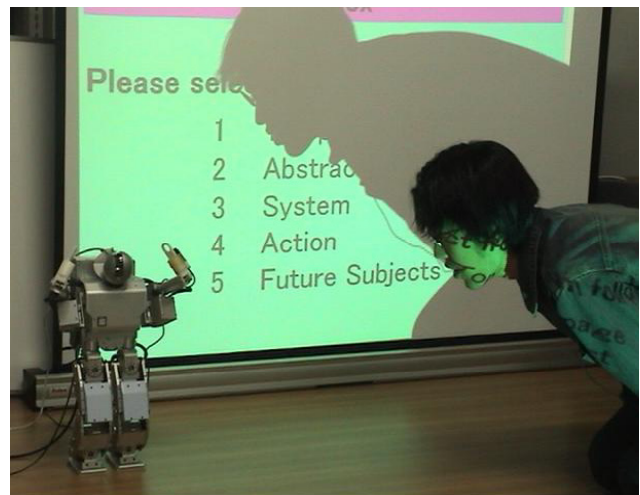The "**`type=`**" argument supports 22 kinds of emotion which are defined by the OCC model.

Figure 4   Simple speech dialog function

3   HUMANOID ROBOT PRESENTATION

3.1   Humanoid Robot

Many humanoid robots have been designed and developed, especially in Japan. We have used the miniature humanoid robot as the presenter, named "HOAP-1". The HOAP-1 (Humanoid for Open Architecture Platform) is the commercial robot that is made by FUJITSU Automation Ltd. It is designed for research and development of robot technologies.
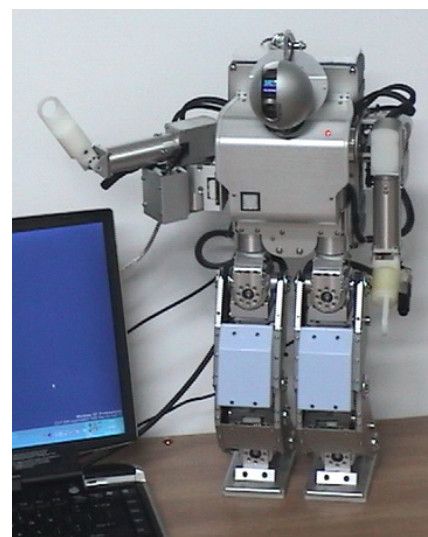

Figure 5   HOAP-1 (FUJITSU) - 483[mm] tall and weigh 5.9[kg]. It has 20 DOF in total. Each leg has 6 DOF and each arm has 4 DOF. A moving CCD camera is attached as the head.

The HOAP-1 is shown in Figure 5. It stands 483[mm] tall and weighs 5.9[kg]. The joint mobility of the HOAP-1 is 20 DOF in total. Each leg has 6 DOF, and it

can walk and turn on two legs. Each arm has 4 DOF. Each hand doesn't have any mechanisms and it can grab nothing. A moving CCD camera is attached as the head of the HOAP-1. This camera isn't a HOAP option. We have attached a small laser pointer on the left hand for pointing the screen, too.
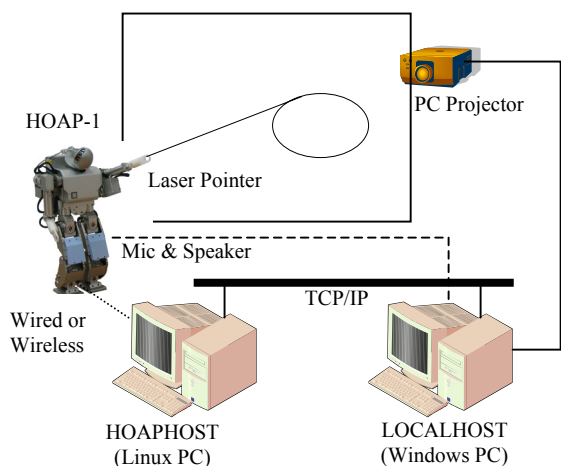
3.2 System Configuration

Figure 6   System configuration

Figure 6 shows the system configuration of our prototype system. It mainly consists of following components.

- humanoid robot HOAP-1
- HOAPHOST (Linux PC)
- LOCALHOST (Windows PC)
- LCD projector and the vertical large screen.

The MPML script is executed on the LOCALHOST (Windows PC). The HOAPHOST (Linux PC) controls the humanoid robot HOAP-1 via wired or wireless connection. Both the HOAPHOST and the LOCALHOST are connected with TCP/IP network.

The MPML script is interpreted on the LOCALHOST, and send control commands for the robot to the HOAPHOST. The LCD projector is attached to the LOCALHOST, then its dispay image is projected on the screen. The HOAP-1 gives a presentation with pointing at the screen. A microphone and a speaker is also connected with the LOCALHOST.

3.3 Humanoid Robot Control

The MPML families have some versions and dialects. There are two kinds of implementations in the MPML families. One is a batch conversion type. Another is an "on the fly" conversion type using the XSL stylesheet technique.

We applied the MPML ver2.0e [8] for the humanoid robot control. The postfix 'e' is an abbreviation of emotion. The MPML 2.0e is a batch conversion type.

We have also developed the variant of the MPML for presentations in 3D space, called the "MPML-VR" (MPML for Virtual Reality) [5]. It can control a 3D character in 3D virtual space. However it uses the VRML techniques, and it has no physical body.
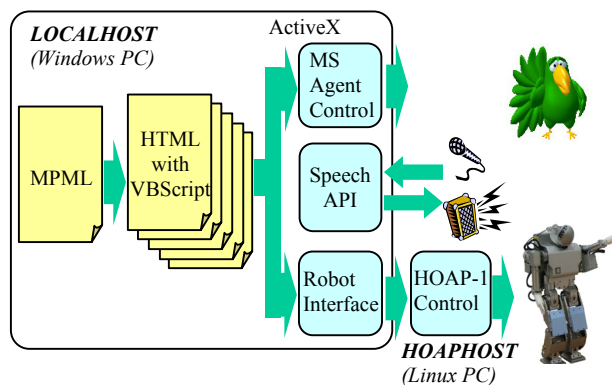

Figure 7   MPML diagram

The MPML diagram is shown in Figure 7.

First, the MPML script is converted to the HTML script with VBScript (or JavaScript). Each part of the MPML script surrounded by the **<page>** and **</page>** tags is converted into an independent HTML file. The single MPML file generates some HTML files. When the HTML file is opened on the Internet Explorer, the VBScript calls up ActiveX components.

We have developed the ActiveX component for the robot interface. We have designed that it has the same interface with the Microsoft agent control. Then, it can select the animated character agent and the humanoid robot at the user's environment. It doesn't need to modify the MPML source codes. The single MPML script can give both animated character presentation on 2D screen and humanoid robot presentation in 3D space.

Whenever a new HTML page including the agent control is opened, the robot interface component establishes TCP/IP connection with the HOAP-1 control on the HOAPHOST. And after the presentation of the page is finished, it closes the connection, since the HTML file is page-based.

The robot interface send control commands through TCP/IP connection to the HOAP-1 control on the HOAPHOST. The HOAP-1 control interprets the commands, and controls the robot. The control commands is independent of the robot.

Figure 8 shows the examples of actions.

The **<play>** tag specifies the pre-defined action of the humanoid robot, since it is difficult for the end-user to

(a) Happy   (b) Disappoint
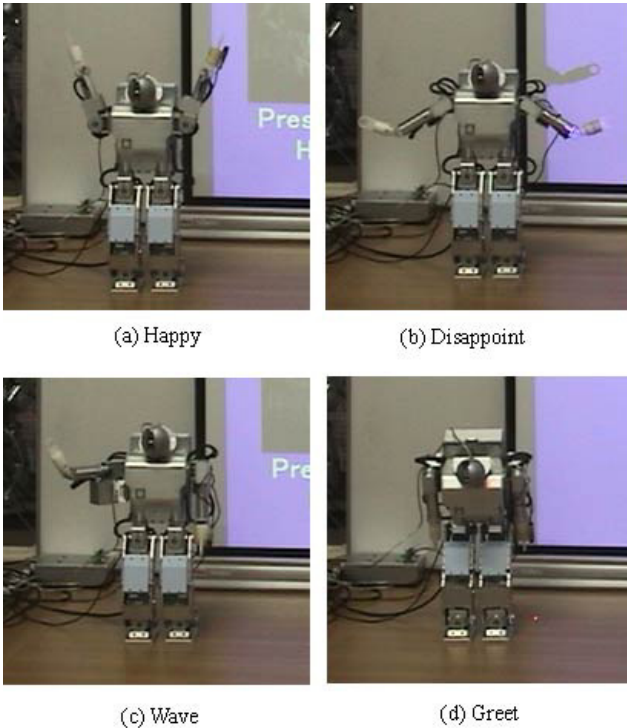
(c) Wave   (d) Greet

Figure 8   The examples of actions

control the transition of all joint angles. The two-legged humanoid robot falls easily down and breaks when the center of gravity of the body deviates.

3.4 Pointing

On 2D presentation, the `<move>` tag moves the character to the position on the vertical screen. However, the humanoid robot can't move vertically. Then, in 3D presentation, we translate the `<move>` tag into the pointing action. The robot points out the position on the screen with the laser pointer. In this prototype system, it isn't support for the robot to move to the arbitrary position, although the robot can walk by the `<play>` action.

The basic pointing action is following.
1. Turn to the screen.
2. Raise the left arm for pointing the position on the screen.
3. Light on the laser pointer.
4. Swing the arm and drawing an oval or a line.

After the presentation of each HTML page is finished, the robot automatically returns to the original posture and faces front to the user. The speed of the turn movement of our robot is modest. It takes considerable seconds to turn to the screen, although the animated character jump to any positions in split seconds. The speed of swinging arm is rapid enough. We have good performance for drawing the oval on the screen like a human.
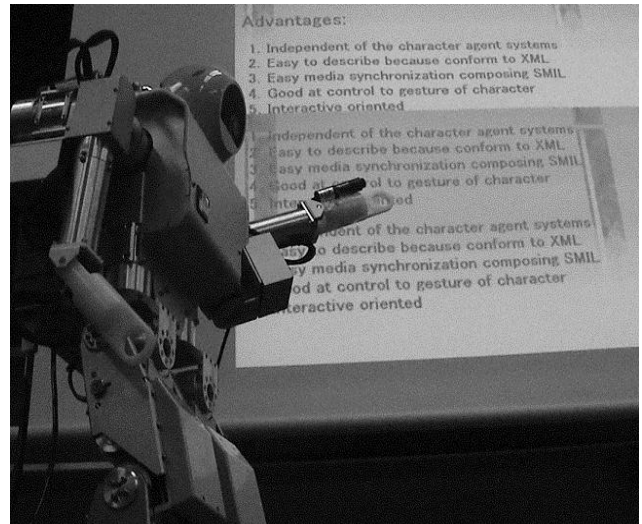

Figure 9   Pointing – The small laser pointer is attached on the left hand, and draws an oval on the screen.

In our prototype system, we calibrate the position and the direction of the robot before the presentation. After the presentation started, motor encoders indicate the position and the direction of the robot. Because the position and the direction error are accumulated, we may require any dynamic compensation methods, for examples, an azimuth sensor or a computer vision feedback system.

4   EVALUATION AND DISCUSSION

A single MPML program can generate both animated character presentation on 2D screen and humanoid robot presentation in 3D space. After we show two presentations, we carried out a questionnaire survey and got the answers from 9 subjects.

4.1 Humanoid robot presentation

Humanoid robot as a presenter

Many of the subjects have agreed about the humanoid robot in 3D space attracting user's attention. However it isn't good that user's attention faces not the presentation but the robot too strongly. When it wants to face user's attention to the screen, it is better restrain the movement of the robot to a minimum. The laser pointer is also useful to move user's attention to the screen.

Pointing with the laser pointer

All subjects have agreed for the pointing action with the laser pointer to be effective. In this prototype system, the robot draws the oval on an important word or an object. It helps for the user to understand the contents of the presentation. But it causes a counter result when the laser pointer points out a different place from the topic.

Emotional expression

In this experiment, we showed two emotional expressions to the subjects, "happy" and "disappoint". It intends to show as non-verbal information a "good thing" and a "bad thing", respectively. Some subjects have recognized the emotion expression of "happy". Only one subject has recognized "disappoint", but he can't recognize "happy". It is difficult for the robot without a facial expression to express the emotion by the gesture only.

4.2 Animated character presentation on 2D screen vs. humanoid robot presentation in 3D space

As a result judged synthetically, three subjects stood up for the robot presentation in 3D space. They have pointed out below as the reason.

· The presentation space becomes stereoscopic.
· Since it is the real world, mental distance with the presenter becomes closely.
· The robot makes the presentation less boring. It is very impressive.

On the contrary, six subjects chose the animated character presentation in 2D space.

· The turn movement between the user and the screen is very slow. Then it divides a series of presentations into some little fragments
· The animated character has high affinity with the display contents since they are in the same virtual world.
· The animated character presentation with the Microsoft agent has excellent quality, whereas the robot presentation hasn't been refined properly yet.
· The robot presentation is attractive. Since it interrupts for the user to concentrate on the presentation, the animated character is better as the presenter.

Many subjects pointed out that the speed of the turning action, for pointing the screen and return to face the user, makes evaluation of the robot presentation improper. It is required to optimize the turn movement. It may be important to change the direction of the robot face, although our robot can't turn the face.

Though many problems became obvious in the comparative experiment with the fine animated character, the favourable comments were also contributed.

5 CONCLUSION

The purpose of the MPML is to provide an easy presentation environment with the character agent. We have combined the multimodal presentation markup language MPML with the humanoid robot. It enables the end-user without programming skills to generate easily his/her own multimodal presentations with the humanoid robot presenter. Since it is designed for the humanoid robot has the same control interface with the animated virtual character, the user can choose either the humanoid robot presentation in 3D space or the animated character presentation on 2D display freely.

A physical robot, especially two-legged humanoid robot, gives strong impression that is different from the animated virtual character on 2D display. However, the robot has some physical restrictions. And the robot presentation hasn't been refined properly yet. We have a plan to apply this system to other humanoid robots.

References

[1] Y. Arafa, K. Kamyab, and E. Mamdani: "Towards a Unified Scripting Language: Lessons Learned from Developing CML and AML", *Life-Like Characters (H. Prendinger and M. Ishizuka eds.)*, Springer, pp. 39-63, 2004

[2] B.D. Carolis, C. Pelachaud, I. Poggi, and M. Steedman: "APML, a Markup Language for Believable Behavior Generation", *Life-Like Characters (H. Prendinger and M. Ishizuka eds.)*, Springer, pp. 65-85, 2004

[3] M. Ishizuka, T. Tsutsui, S. Saeyor, H. Dohi, Y. Zong and H. Prendinger: "MPML: A Multimodal Presentation Markup Language with Character Agent Control Functions", *Proc. Agents2000 Workshop 7 on Achieving Human-like Behavior in Interactive Animated Agents*, pp.51-54, 2000

[4] A. Marriott: "VHML – Virtual Human Markup Language," (*Online*), http://www.interface.computing.edu.au /document/VHML

[5] N.Okazaki, S.Aya, S. Saeyor, and M.Ishizuka: "A Multimodal Presentation Markup Language MPML-VR for a 3D Virtual Space", *Workshop Proc. (CD-ROM) on Virtual Conversational Characters: Applications, Methods, and Research Challenges (in conjunction with HF2002 and OZCHI2002),* 4 pages, 2002

[6] S.Sugano, K.Shibuya: "Anthropomorphic Robots for Nonverbal Communication", *Journal of the Robotics Society of Japan*, Vol.15, No.7, pp.975-978, 1997

[7] J.Yamato, R. Brooks, K. Shinozawa, and F. Naya: "Human-Robot Dynamic Social Interaction", *NTT Technical Review*, Vol. 1, No.6, pp.37-43, 2003

[8] Y. Zong, H. Dohi and M. Ishizuka: "Multimodal Presentation Markup Language MPML with Emotion Expression Functions Attached", *Proc. 2000 Int'l Symp. on Multimedia Software Engineering (IEEE Computer Soc.),* pp.359-365, 2000

[9] Microsoft Agent Home Page: http://www.microsoft.com/msagent

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

[11] TVML Home Page: http://www.tvml.tv