Humanoid Robot Presentation through Multimodal Presentation Markup Language MPML-HR

Kazutaka Kushida, Yoshitaka Nishimura,
Hiroshi Dohi, Mitsuru Ishizuka
School of Information Science and Technology,
University of Tokyo
7-3-1, Hongo, Bunkyo-ku
Tokyo 113-8656, Japan
{ kushida, nisshi, dohi, ishizuka}
@miv.t.u-tokyo.ac.jp

Johane Takeuchi, Hiroshi Tsujino
Honda Research Institute Japan Co., Ltd.
8-1, Honcho, Wako-shi
Saitama 351-0188, Japan
{ johane.takeuchi, tsujino}
@jp.honda-ri.com

ABSTRACT

Multi-modality is becoming important to improve the understandability and friendliness in human-computer interfaces and media contents. We have developed a Multimodal Presentation Markup Language, called MPML. MPML is a simple script language to make multi-modal presentation contents using animated characters for presenters. As MPML is based on XML, anyone can write MPML contents easily. In conventional MPML we used animated characters as presenters. The humanoid robot is more life-like than the animated characters, because it can move and act in a real space, and look at the audience. So humanoid robots can be a more attractive presenter. In this paper, we describe a new version of MPML called MPML-HR, where HR stands for Humanoid Robots. Presentation contents written in MPML-HR are performed by a biped humanoid robot instead of an animated character in 2D display. Anyone can control humanoid robots and make them do presentation using MPML-HR without expertise of humanoid robots. In this paper, we show the development of MPML-HR and its application to the Honda humanoid robot, ASIMO.

Keywords

humanoid robot, multi-modal presentation, life-like agent

1. INTRODUCTION

Not only documents but also multi-modal contents including visual and auditory information are used in media and interface fields these days. We can see an animated character who acts and speaks to serve as an attendant to a guest on the web of a shopping site. A multi-modal CG character sometimes appears as an assistant in TV program.

In order to make multi-modal contents using life-like characters easily, several languages have been developed [1]. VHML (Virtual Human Markup Language) [2] and CML (Character Markup Language) [3] are both scripting and representation languages for the animation of character agents. APML (Affective Presentation Markup Language) [4] is a script language to make facial expressions . TVML (TV Program Making Language) [5] and our MPML (Multimodal Presentation Markup Language) [6] are scripting languages for presentations with character agents, while the former one is especially for TV programs.

As MPML is a medium-level scripting language, anyone can write multi-modal presentation with life-like characters. Before we used an animated character as a presenter but humanoid robot can be a more attractive presenter because it can move in real world.

In regard to robots, the recent development of humanoid robots is remarkable. Many humanoid robots are expected to play roles in entertainment, security, rescue, care, guiding and so on. However, it is necessary to know the specific languages to control them. Due to this constraint it is difficult for non-professional users to control them. To address this problem and effectiveness of using humanoid robots as presenter we have developed a new version of MPML called MPML-HR for humanoid robots. Using MPML-HR it is easy to write and generate a series of humanoid robot behaviors. The presentation by the humanoid robot gives a different impression from that by the character agent; it is much more life-like because, for example, it can move in real space and look at audiences.

Conventional MPML uses an animated character appears only on the display of PC or mobile phone. Because MPML-HR uses humanoid robots as presenter, we can extend presentation contents explaining actual objects(for example, new goods and digital cameras and cars).

In this paper, we describe the development of MPML-HR, and its application to the Honda humanoid robot, ASIMO. There are several studies related to the humanoid robots, including ours [10, 11, 12].

AAMAS'05, July 25-29, 2005, Utrecht, Netherlands.

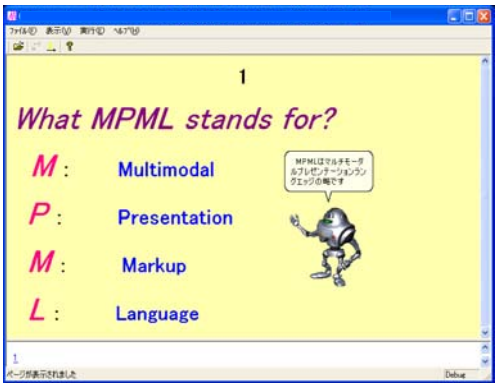

Figure 1: Presentation image of conventional MPML

2. MULTIMODAL PRESENTATION MARKUP LANGUAGE MPML

MPML has been developed to allow many people to easily write attractive multi-modal presentations using life-like agents. Figure 1 is an example view of conventional MPML presentation.

2.1 The features of MPML

Though several versions of MPML have been developed, this subsection shows common features. The features of MPML are following.

- Independent of particular systems
  MPML is a medium level description language that does not depend on particular browsers or agent systems. The multi-modal contents made by MPML can be performed in various environments, when a converter for the environment and the agent system to be used are prepared.
- Character control function
  MPML has many functions controlling a character agent: position, movements, gestures and the emotion of the agent.
- Easy description
  MPML is based on XML. Unlike programming languages it is easy to write multi-modal contents with MPML using its tags whose meanings are understandable not only to a computer but also to a human.
- Media synchronization
  MPML has a function for media synchronization based on SMIL.
- Meta-data description
  Meta-data such as the author of the content, time of presentation and information for copyright can be added to the MPML texts.

2.2 MPML family

There are several versions of MPML in the MPML family. MPML Ver.1.0 is the first version of MPML developed in 1998. MPML Ver.2.0, the second version of MPML, supports plural agents. In MPML Ver.2.0e [8], an extension of

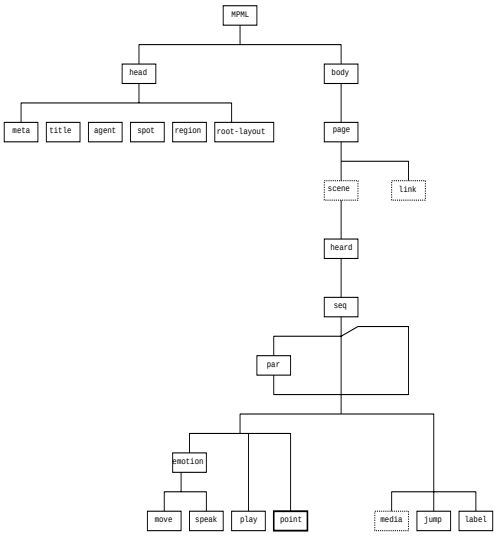

Figure 2: Specification tree of MPML-HR

Ver.2.0, an emotion tag was introduced to allow easy description of emotional behavior and speech synthesis of the agents. MPML-HR was designed based on this Ver.2.0e. In MPML Ver.3.0, some extensions including a graphical editor have been incorporated. MPML usually uses Microsoft agent as presenter, but the version using different agents also has been developed. MPML-VR [9] stands for MPML for Virtual Reality. It can control a 3D character in 3D virtual space, and uses VRML techniques. MPML-flash is the flash version of MPML, where multi-modal contents are synthesized in the Flash format in a server and sent to a client. As a result, a client computer without a character agent functionality can generate the contents. MPML-mobile allows that the multi-modal contents with the character agents to run on cellular phones, though at present the contents do not include speech modality. In MPML-mobile system, texts appear on the display instead of synthesized speech.

2.3 Tags

MPML has several tags for the description of multi-modal presentation contents. As these tags are almost same as MPML-HR tags (see next section).

3. MPML-HR: MULTIMODAL PRESENTATION MARKUP LANGUAGE FOR HUMANOID ROBOT

3.1 MPML-HR tags

Figure 2 shows the specification tree of MPML-HR which is used to describe not only the behaviors of the humanoid robot and character agents on a display, but also the control of presentation materials. Basic tags of MPML-HR are summarized following. These tags are usable for conventional MPML except for < **point** > tag.

①Head tag

The data surrounded by the < **head** > tag indicates the management information of the presentation contents.

②Title tag

The < **title** > tag shows the title of the presentation.

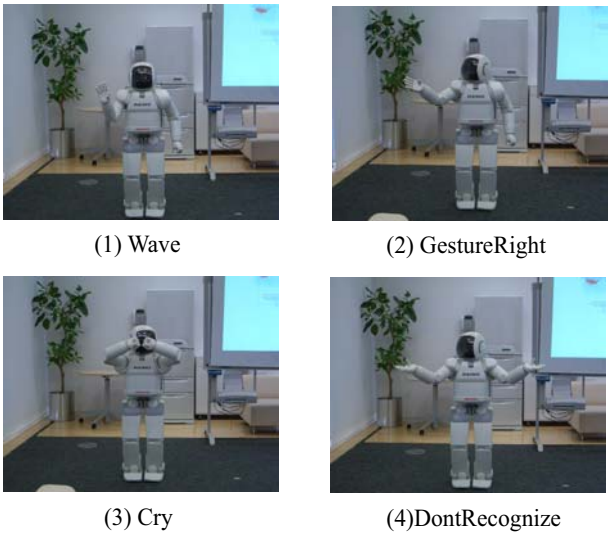

Figure 3: The examples of "play" actions

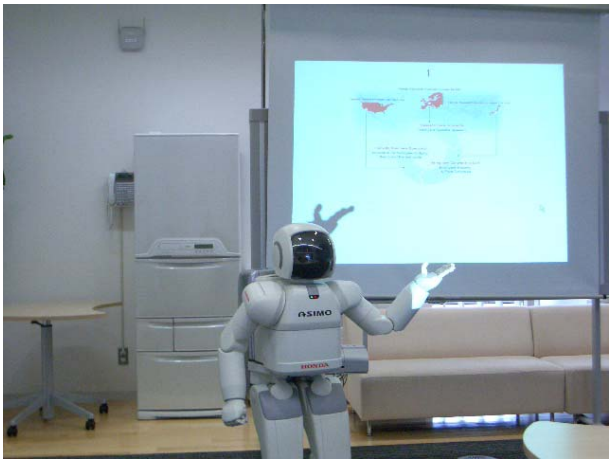

Figure 4: The pointing action

③**Agent tag**
The agent performs the presentation which is specified by the < **agent** > tag. The MSAgent (Peedy, Genie, Merlin and so on) or a humanoid robot, Honda's ASIMO in our case, are specified by {character=""} in this tag.

④**Body tag**
The contents which surrounded by the < **body** > tag are the main part which constitutes the presentation.

⑤**Page tag**
The < **page** > tag indicates a page break in the presentation. One presentation segment usually contains one PPT page or HTML page as presentation screen material. An MPML-HR content, like other MPML cases, may include more than one page. In conventional MPML, HTML pages are usually used and PPT pages cannot be used. In MPML-HR, not only HTML pages but also PPT pages can be used for explanation targets.

The tags often used for the performance of the agents are summarized as follows.

⑥**Play tag**
The < **play** > tag invokes a specified action. Each agent has some pre-defined actions. The action is specified by {act=""}. ASIMO has many pre-defined actions such as wave, gesture right, cry, don't recognize, etc. People without expertise can make his/her presentation contents easily by using the < **play** > tag, because he/she only describes this tag to invoke a specified action, and don't need to care about the detailed control of the action. Figure 3 shows the examples of the ASIMO's actions.

⑦**Move tag**
The < **move** > tag indicates the movement of the humanoid robot to the specific position. In case of conventional MPML, this tag has only two attributes "x" and "y". The position where an animated character go is specified by them. In case of MPML-HR, this tag has three attributes "x" and "y" and "z". The position is specified by {x=""} and {y=""}, where the (x,y) position is x[m] or y[m] far from the origin and the direction of a humanoid robot is specified by {z=""} .

⑧**Speak tag**
The humanoid robot speaks the text sentence surrounded by the < **speak** > tag using a speech synthesizer.

⑨**Emotion tag**
The < **speak** > tag and the < **move** > tag can be surrounded by the < **emotion** > tag. The pitch and tempo of synthesizing speech are changed by this tag according to the emotion. In addition, some appropriate gestures are introduced before or after the speech, if suitable, to manifest the emotion. In MPML Ver.2.0e, the speed of the movement is changed by this tag; however, in MPML-HR, this functionality is excluded. There are several models of emotion such as evaluation-activation [14]. In our system the emotion is specified by {types=""}. Based on the OCC model, the famous emotion model, 22 types of emotion such as 'happy-for', 'worried', 'anger', etc. are prepared in the current implementation.

⑩**Point tag**
The < **point** > is a new tag introduced in MPML-HR. In conventional MPML using a 2D character agent, the agent can point to the target on the screen by moving to the target directly. Here the < **move** > tag is used to point to the specified position on the screen. In the humanoid robot, however, he doesn't move to the position on the screen but move real world. To point the specified position on the screen, the humanoid robot points on the screen with his hand, a stick or a laser pointer as in human behavior. To solve this problem, the < **point** > is added to MPML-HR. The < **point** > is used with the attribute of {x="" y=""}. The robot can point the (x,y) position on the screen using this tag. In the current implementation, the robot points the screen not with a laser pointer or a stick, but with his left or right hand. The robot moves to the screen and turns 20 degrees inside and raises his hand. The robot can point 1 of 6 areas on the screen. If the point operation is executed, the robot moves to one side (right side or left side) of the screen, and turns 20 degrees inside and then points the high area, middle area, or low area with his hand. If the target point is on the right half of the screen, the robot moves to the right side and points the area with his left hand. If the contrary, he moves to the left side and points the area with his right hand. This system also supports the robot moving to an arbitrary position by the < **move** > tag.

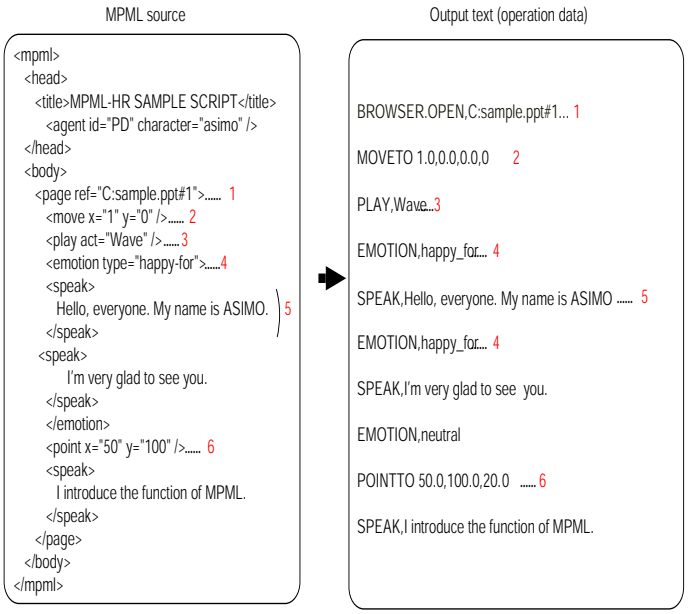

Figure 5: An example of MPML-HR

A flow of pointing action execution is following.

1. approach to one side of the screen (right side or left side)
2. Turn his body 20 degrees to the screen.
3. Raise the right or left hand, and point the high, middle or low area of the screen

3.2 A simple example of the MPML-HR presentation

The left side of Figure 5 shows an example of the MPML-HR description. As MPML is based on XML format, people can write MPML source like writing html source. From Figure 5, one can see that the MPML description is simple and understandable. An MPML source file is converted to an operation data file by an MPML-HR parser program. The right side of Figure 5 shows a converted operation data file.

The explanation of this example (MPML source and Operation data) is following.

1
This part defines the page which is explained by a humanoid robot. The page is projected on the screen by the LCD projector through the Windows PC. In this sample, the first page of "sample.ppt" is projected on the screen. The other format files also can be projected on the screen, for example html files.

2
The move operation is described. ASIMO moves to 1[m] on the x-axis, 0[m] on the y-axis. In this sample, only two attributes are written, but another attribute "z[deg]" is prepared. "z" is the direction of a humanoid robot. A humanoid robot move to the point designated by "x" and "y" and turn to the angle designated by "z". If "z" is not written in the script, a humanoid robot does not turn.

Figure 6: ASIMO(Honda).

3
The play operation is described. ASIMO waves to audience. Many action patterns are prepared, for example, "Greet", "GestureRight".

4
The emotion description is here. The volume and the speed of speech are defined every emotion. The speech which is surrounded by the emotion tag is modified according to that.

5
The speak operation is described. ASIMO speaks, "Hello, everyone. My name is ASIMO." As this speak operation is surrounded by the emotion tag, ASIMO happily speaks.

6
The point operation is described. ASIMO moves to the left side of the screen, turns his body 20 degrees to the inside, and raises his right hand. This is equal to the description "move" and "play act="GestureRight"".

4. A SYSTEM OF MPML-HR PRESENTATION

4.1 Agent(Humanoid robot)

Several humanoid robots have been designed and developed, especially in Japan. We use a humanoid robot named "ASIMO" as the presenter. The name of "ASIMO" is derived from Advanced Step in Innovative MObility. ASIMO, made by Honda, is a biped walking robot.

The ASIMO is shown in Figure 6. It stands 1200[mm] tall and weights 52[kg]. The joint mobility of the ASIMO is 26 DOF in total. Each leg has 6 DOF, and it can walk and turn on two legs. Each arm has 5 DOF. Each hand has 1 DOF, and he has 500 [g] grips so that he can hold something light. The head has 2 DOF, and he can shake or nod his head. He can walk at 1.6 km per an hour. He has several kinds of movements like dance, greet, guide, nod and so on.

4.2 System Configuration

Figure 7 shows the configuration of our system. It mainly consists of the following components.

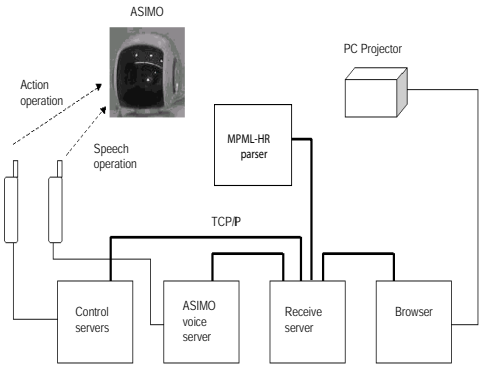

Figure 7: System configuration

- Humanoid robot ASIMO
- Control Servers(1 Linux PC and 1 Windows PC)
- ASIMO Voice Server(Windows PC)
- Receive Server(Windows PC)
- MPML-HR parser(Windows PC)
- Browser(Windows PC)
- LCD projector , speaker and the vertical large screen

The Receive Server, the MPML-HR parser and the Browser can be usable not only on the different PCs but also on the same PC. The MPML script is executed by the MPML-HR parser (Windows PC). The Control Servers (1 Linux PC and 1 Windows PC) control the humanoid robot ASIMO via wireless connection. The Receive Server (Windows PC) receives operation data from the the MPML-HR parser. This Server reads mapping table of operation data and ASIMO commands. Then it sends the equivalent ASIMO commands to the Control Servers, and sends the speak commands to the ASIMO Voice Server (Windows PC). If the new page is opened, the Receive Server also sends the open command to the Browser (Windows PC).

The MPML-HR parser and the Receive Server, the Receive Server and the ASIMO Voice Server, the Receive Server and the Browser, and the Receive Server and the Control Servers are connected via TCP/IP network.

The MPML script is interpreted into operation data by the MPML-HR parser, and the MPML-HR parser sends operation data to the Receive Server. The Receive Server sends commands to the Control Servers or the ASIMO Voice Server or the Browser. The LCD projector is attached to the Windows PC on which the Browser is. The ASIMO performs a presentation by doing designated action, for example, pointing the screen. A speaker is connected to the ASIMO Voice Server.

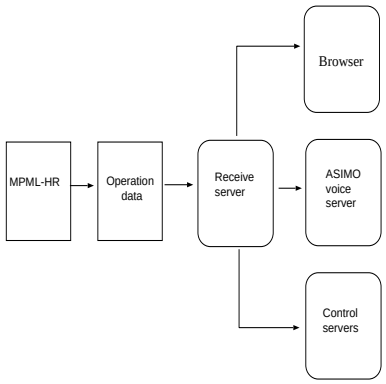

Figure 8: MPML-HR control diagram

4.3 MPML-HR parser

Figure 8 shows the control diagram of the MPML presentation.

A conventional MPML source code using 2D animated Characters is converted to a VBScript code by an MPML converter. Characters are controlled by the VBScript command.

We have developed the MPML-HR parser as the robot interface. In conventional MPML for 2D agent, it calls the ActiveX component of the MSAgent interface. The MPML-HR parser corresponds to the MPML converter and the MSAgent interface.

First, MPML-HR parser opens the MPML-HR source which is based on XML format. And MPML-HR parser parses it and examines whether it coincides with XML format. Then MPML-HR parser sends it to the XSLT processor. The XSLT processor converts a MPML-HR source to operation data whose format can be understandable for the Receive Server according to converting rule described in a style sheet.

MPML-HR parser establishes a TCP/IP connection with the Receive Server. As the Receive Server is connected with the ASIMO Voice Server and the Control Servers and the Browser via TCP/IP, the network path from the robot interface component to the ASIMO control on the Control Servers is established.

The MPML-HR parser reads the operation data made by the XSLT processor and sends them through the TCP/IP connection to the Receive server. The operation data is independent of the robot. The Receive Server interprets the data into control commands referring the mapping table and sends the commands through the TCP/IP connection to the ASIMO control on the Control Servers. The ASIMO control controls the robot.

5. MAKING THE PRESENTATION MORE LIFE-LIKE ...

5.1 Parallel execution of the action command and the speech command

In ASIMO, the action command and the speech command are operated independently, so that 2 threads are generated

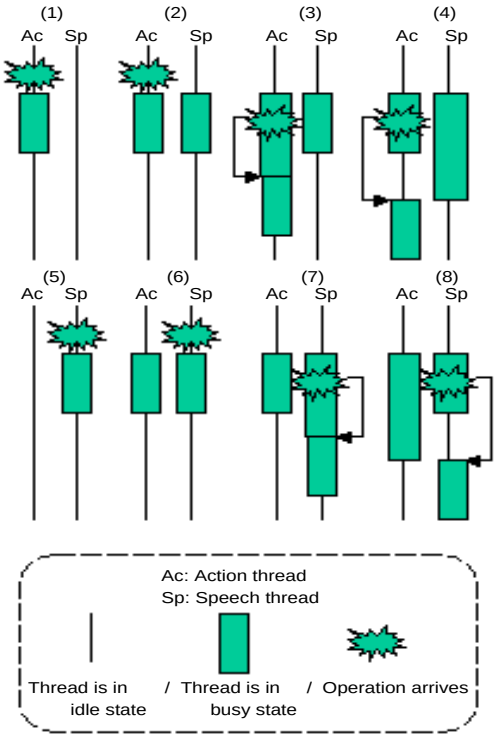

Figure 9: Timing of the action command and the speech command execution

in the Receive Server (Figure 7). One is the action thread and the other is the speech thread.

The behavior of the Receive Server is described below. The Receive Server receives the operation data from the MPML-HR parser, and dispatches the commands to the ASIMO Voice Server (if the command is a speech command) and the Control Servers (if the command is an action command) (section 4.2) and Browser (if the command is to open the specified page). If the Receive Server receives the action command and the action thread is in idle state, the action command is sent to the Control Servers immediately, whether the speech thread is in busy or idle states ((1) or (2) of Figure 9). When the Receive Server sends an action command (or a speech command) to the Control Servers (or the ASIMO Voice Server), the action thread (or the speech thread) turns into busy state from idle state until the end signal is received from the Control Servers. Next, if the Receive Server receives the action command and the action thread is in busy, the action command is sent to the Control Servers after both the action thread and the speech thread become idle state ((3) or (4) of Figure 9). The rest of figure 9 describes the threads behaviors in the other conditions ((5) - (8) of Figure 9).

5.2 Autonomous head motion

In case of conventional MPML, a 2D animated agent works his mouth while he is speaking. But a humanoid robot without mouth doesn't do anything while he is speaking, if he finishes the specified action by < **play** > tag. To make humanoid robot presentation more life-like, the head action

Figure 10: Examples of autonomous head motion

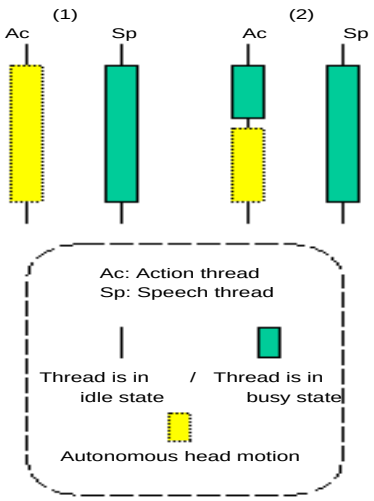

Figure 11: Timing of autonomous head motion execution

command is automatically sent at random from the Receive Server. By moving his head, we expect he could be observed as more life-like agent and audience glanced by him would get a sense of unity. This head action command makes the robot move his head right and left, and up and down. The degree of motion is determined at random within some range. Figure 10 shows this autonomous head motion.

The head action command is sent at the time that the robot is speaking with no action. In the Receive Server, this means that the speech thread is in busy state and the action thread is in idle state. There are two situations in receiving the head action command. One is that the robot is speaking with no action ((1) of Figure 11). The other is that the robot is speaking with the short action but the action ends before the speech ends. In the latter case, the head action command is received after the action ends ((2) of Figure 11).

6. CONCLUSION

We have extended MPML to MPML-HR MPML-HR is the script language for humanoid robot presentation. The MPML-HR description is so simple that anyone can write attractive multi-modal presentation contents using a humanoid robot without expertise of the robot. As MPML-HR description is almost same as conventional MPML, conventional MPML contents are able to be used for MPML-HR contents. Since at present there is not a good tool or programming system for making the humanoid robot presentation, MPML-HR is also effective even for professional users when making long and complicated interactive contents. As a humanoid robot is off the display, MPML-HR can be used to real object presentation(for example, digital camera, car, TV).

To make the robot presentation more life-like, we made the Receive Server have 2 threads to allow parallel execution of the action command and the speech command. Also, autonomous head motion is generated during the speech. Autonomous head motion give a sense of unity to audience glanced by the robot.

In experiments to date, the audience appeared to receive a stronger impression looking at the humanoid robot presentation rather than the 2D animated agent presentation. Although there has been only a small amount of experimental work, the multi-modal presentation by the humanoid robot appears to convey rich information and thus help effective and deep understanding by the audience.

Our future work includes making the robot presentation more life-like, for example, to give the robot autonomous actions or interactive actions. Now pointing action is only to raise the right hand or left hand, so it is difficult to point target accurately. We plan to use an animated agent for pointing action with a humanoid robot.

7. ACKNOWLEDGEMENT

We especially thank Dr. Tatsuhiko Sekiguchi, Mr. Yuji Hasegawa, Mr. Toyotaka Torii, Dr. Mikio Nakano, Dr. Masahiro Kimura, Dr. Kazuhiro Nakadai and Mr. Kentaro Yamada, who are members of Honda Research Institute Japan Co., Ltd., for their helps on the demonstration of MPML system and the operation of ASIMO.

8. REFERENCES

[1] H. Prendinger and M. Ishizuka (eds.) : "Life-like Characters-Tools, Affective Functions, and Applications, Springer-Verlag, 2004."

[2] A. Marriott: "VHML - Virtual Human Markup Language," (Online), http://www.interface.domputing.edu.au/document/VHML

[3] 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.

[4] 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.51-54, 2004.

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

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

[7] H. Prendinger, S. Descomps and M. Ishizuka: "MPML: A Markup Language for Controlling the Behavior of Life-like Characters," Jour. of Visual Languages and Computing, Vol.15, No.2, pp.183-203, 2004.

[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] N. Okazaki, S. Aya, S. Saeyor, and M. Ishizuka: "A Multimodal Presentation Markup Language MML-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.

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

[11] 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

[12] Y. Nozawa, H. Dohi, H. Iba and M. Ishizuka: "Humanoid Robot Presentation Controlled by Multimodal Presentation Markup Language MPML," Proc. 13th IEEE int'1, Workshop on Robot and Human Interactive Communication (RO-MAN2004), No.26, 2004.

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

[14] R. Cowie, E. Douglas-Cowie, S. Savvidou, et al.: "Feeltrace: An instrument for recording perceived emotion in real time," In ISCA Workshop on Speech and Emotion (Northern Ireland, 2000), pp.19-24.