MPML for Describing Multimodal Contents with Lifelike Agents

Mitsuru Ishizuka
Dept. of Information & Communication Eng.
& Dept. of Creative Informatics
School of Information Science and Technology
University of Tokyo

Progress in Lifelike Embodied Agents

- Research Activities from approx. 1990 at
  - DFKI, USC/ISI, CMU, NCSU, Stanford, MIT, Univ. of Rome, Curtin Univ. of Tech., Microsoft, etc.,
  - and Univ. of Tokyo

have been showing the feasibility and positive effect as new multimodal media and new educational media.

- Necessary media components are becoming available.

Some Cognitive Backgrounds

⌘ Non-verbal Communication

by *Albert Mehrabian*

via Language (flat sentence)	7%
via Speech with tone and intonation	38%
via Facial expression and Gesture	55%

⌘ The Media Equation

by *B. Reeves, C. Nass*

Media = Real Life

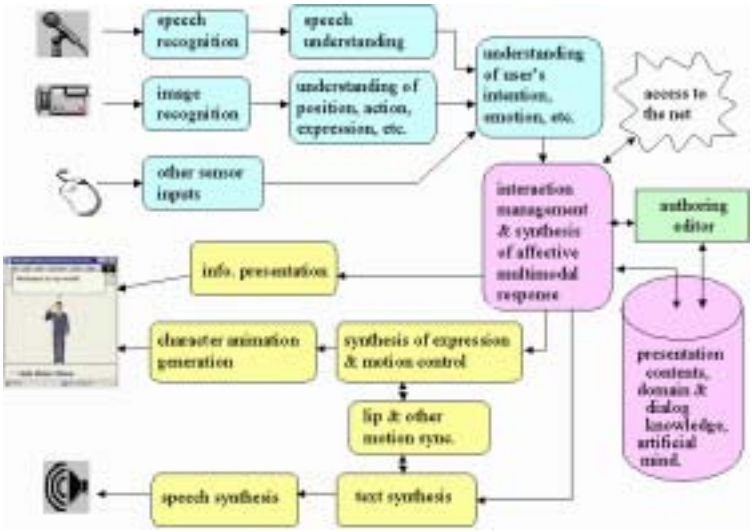
⌘ The Persona Effect

The presence of a lifelike character even one that is not expressive - can have a strong positive effect on student's perception of their learning experience.

Dimensions:

motivation, entertainment, helpfulness, …

Many Component Technologies are necessary to build a system.



Rich & Cool Multimodal Media not only for everyone, but also by everyone

Need for XML-based Description Language

- **MPML** (Multimodal Presentation Markup Language
- **VHML** (Virtual Human Markup Language)
- **CML/AML, APML, RRL-NECA, BEAT, …**
- **HumanML** (Human Markup Language)

Some Competitors

- **VoiceXML**
- **Web3D (Shockwave3D, …….)**
- **MS Narrator**

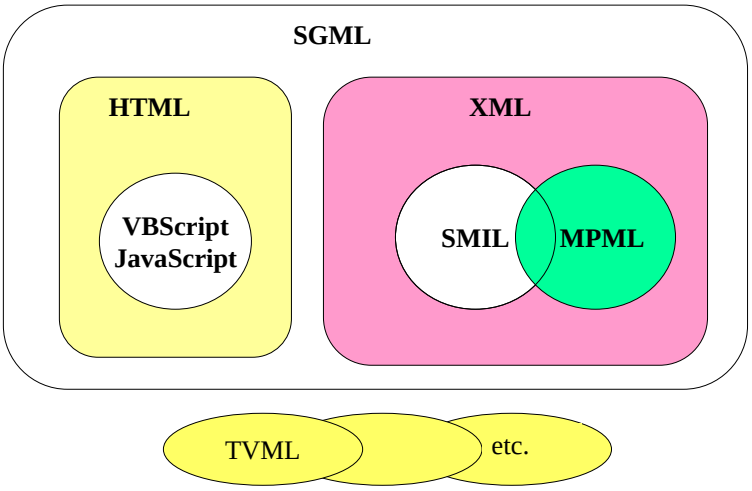
MPML concept

(Multimodal Presentation Markup Language)

- **Multimodal Presentation Anytime, Anyplace through the network (even to mobile).**
- **Allows Anyone (ordinary people) to write effective/attractive Multimodal Presentation Contents easily.**

- **Serves as an extensible center integrating many advanced functional modules.**

MPML as a Markup Language conformed to XML

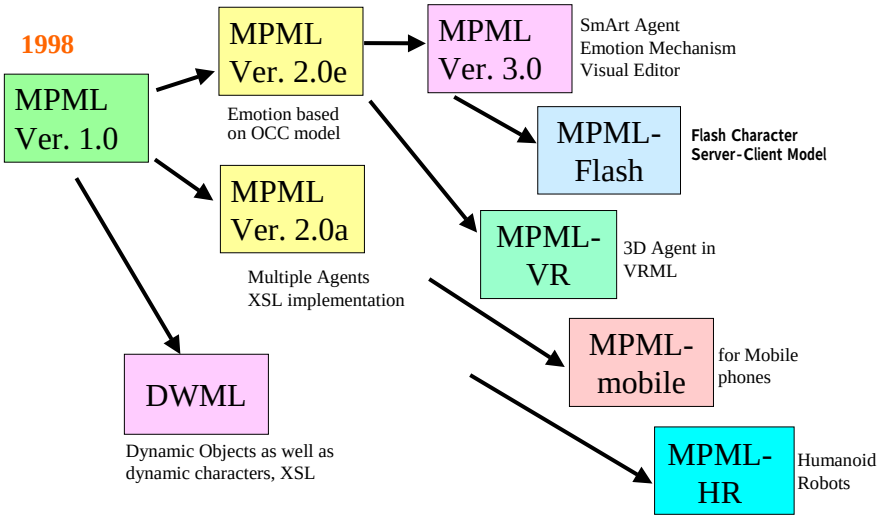


⌘ SMIL: Synchronized Multimedia Integration Language

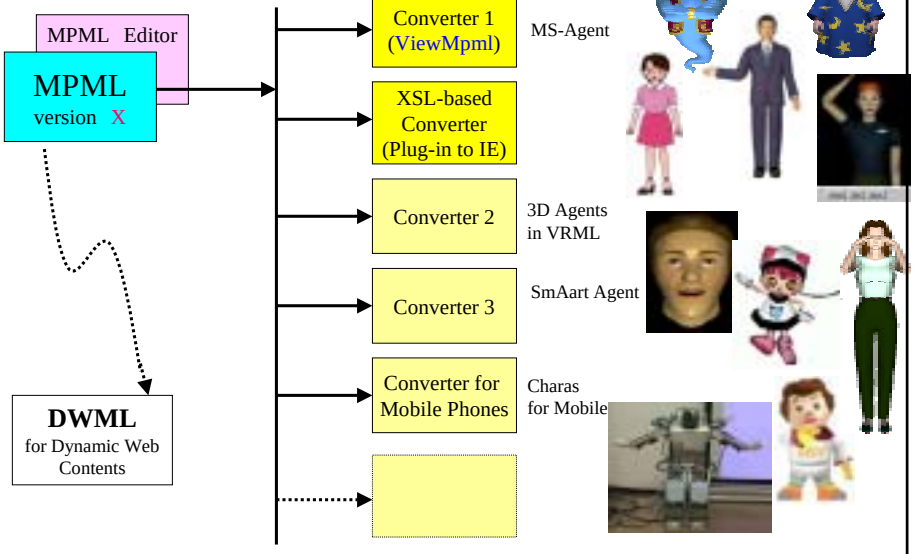
A simple example of *MPML* script

```
<mpml>
<head>
  <spot id="spot1" location="200,260" />
  <agent id="simasan" system="MSAgent"  character="simasan"
      voice="LH"  agreeableness="50" activity="50" spot="spot1" />
</head>
 <body>
    <seq>
      <scene agents="simasan">
        <page ref="page0.html">
          <act agent="simasan" act="greet"/>
          <speak agent="SmArt1">
          <emotion assign="simasan:happy+"/>
            Hello!  My name is Sima.  Welcome to our Web.
          </speak>
        </page> </scene> </seq>
</body>
</mpml>
```

History of MPML

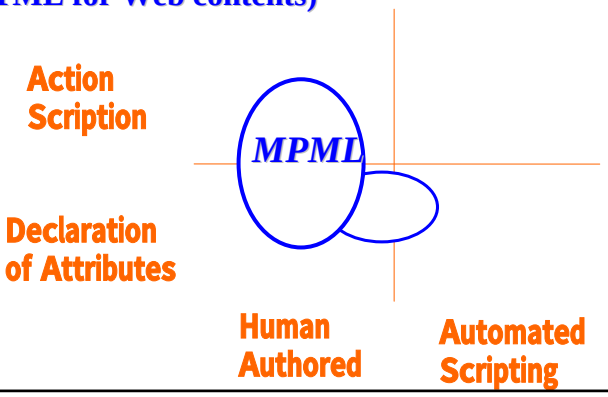


MPML Play -- several ways



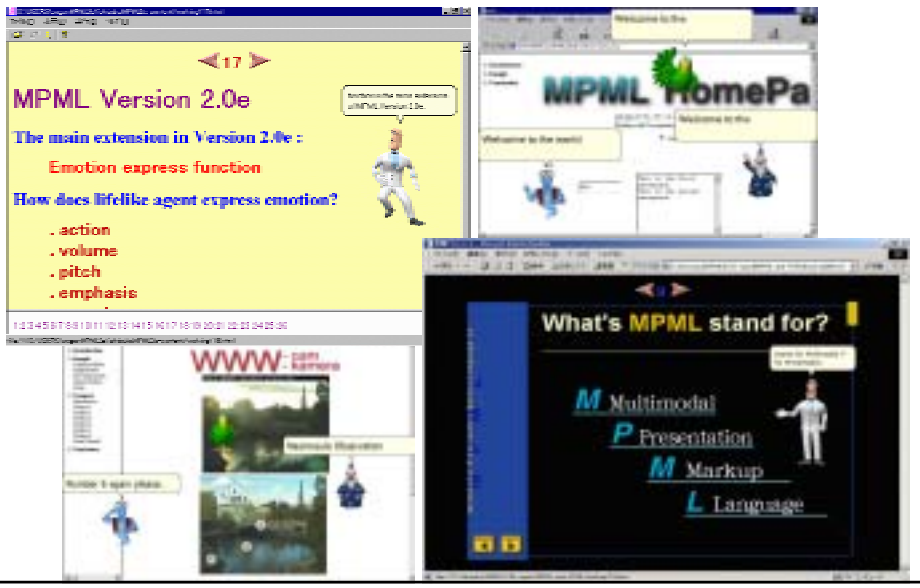
MPML's position in the taxonomy of description languages

- Medium level
  - dialogue management based on finite state machine and a memory mechanism.
- Easy authoring for ordinary people in language level (like HTML for Web contents)



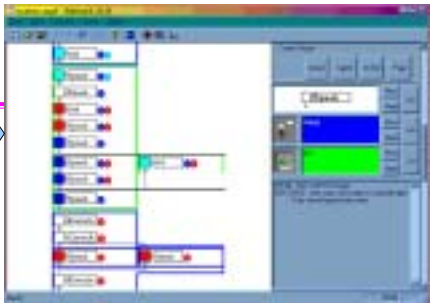
MPML 2.0 featuring

Full Presentation with Emotional Expressions



MPML3.0

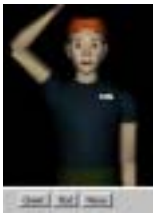
Graphical Editor

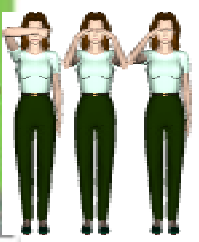


with SmArt Agents

MPML-VR

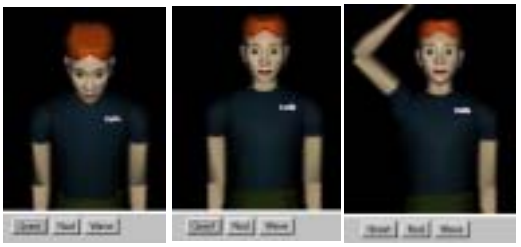
-- Presentation in 3D VRML Space





3D Agents in VRML space

Andy and Aya



MPML-HR (humanoid robot) version

MPML-HR for Honda's ASIMO



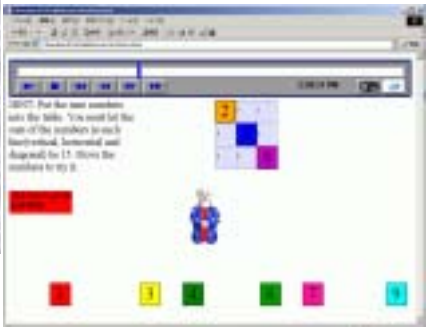

MPML-mobile 0.5 for J-Phone (J-Sky) and 1.0 for DoCoMo's i-mode

In Cooperation with Hottolink Inc.

DWML

Dynamic Web Markup Language

- Animation control not only for character agents, but also for all objects.



Features required for Affective Lifelike Agents

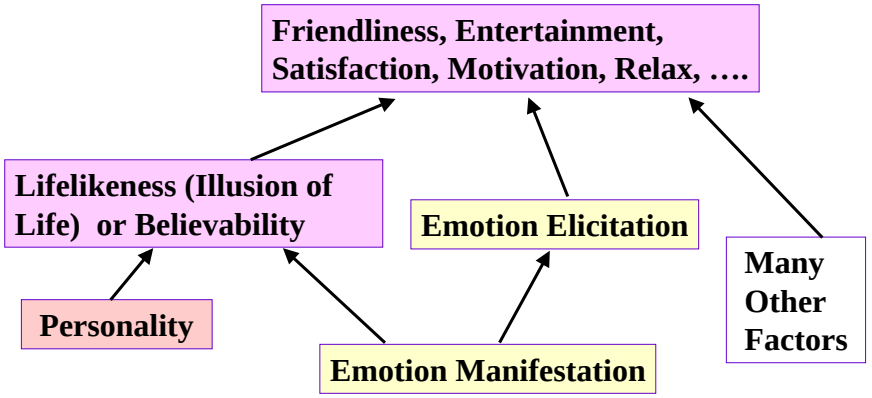
Embodiment

- Synthetic Bodies
- Emotional Facial Display
- Communicative Gestures
- Posture
- Affective Voice

Artificial Emotional Mind

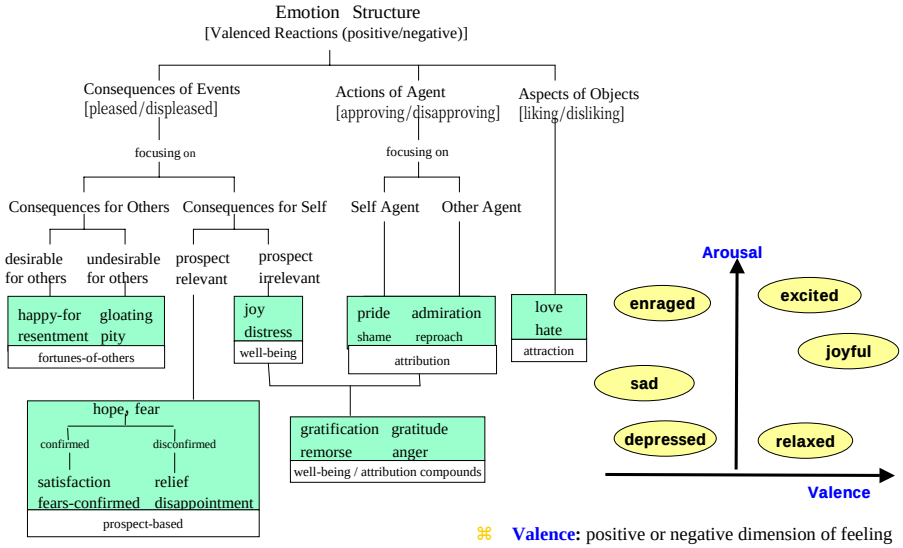
- Affective-based Responses
- Personality
- Response adjusted to Social Context
  - Social role awareness
- Adaptive Behavior
  - Social intelligence

More Affective and Social
⇨ Workable Model of Emotion/Personality



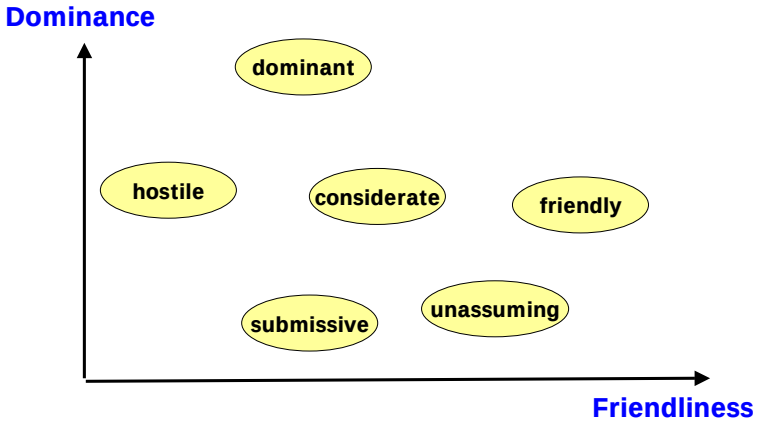
Affect expression is also an important feedback channel in communication. For example, when your interlocutor frowns, you know something is wrong in the conversation.

OCC (22 emotions) and Lang's 2-dimensional Emotion Models



- **Valence:** positive or negative dimension of feeling
- **Arousal:** degree of intensity of emotional response

McCrae and Costa's 2-dimensional Personality Model (89)



- Dominance: individual's disposition to control
- Friendliness: tendency to be warm and sympathetic

Scripting Emotion in *MPML2.0e*

```
<mpml>
<head>
  <title> MPML Presentation </title>
</head>
<body>
<page id='first' ref='self_intro.html'>
<emotion type='happy-for'>
  <speak>
   I am Mitsu Ishizuka from the Univ. of Tokyo.
  </speak>
</emotion>
</page>
</body>
</mpml>
```

'happy-for'

Emotion and Voice Parameters

Emotion	*Fear*	*Anger*	*Sadness*	*Happiness*	*Disgust*
Speech rate	much faster	slightly faster	slightly slower	faster or slower	very much slower
Pitch average	very much higher	very much higher	slightly lower	much higher	very much lower
Pitch range	much wider	much wider	slightly narrower	much wider	slightly wider
Intensity	normal	higher	lower	higher	lower
Pitch changes	normal	abrupt on stressed syllables	downward inflections	smooth upward inflections	wide downward terminal inflections

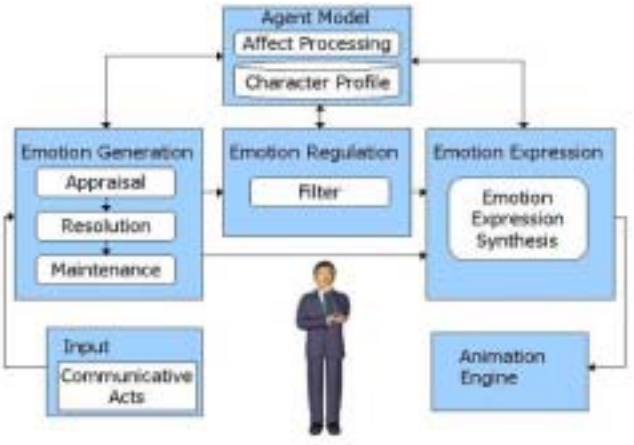
(The emotion of "grief" is omitted.)

Emotion	*Fear*	*Anger*	*Sadness*	*Happiness*	*Disgust*
Speech rate	+30	+10	-10	+20/-20	-40
Average pitch	+40	+40	-10	+30	-40
Loudness	—	+6	-2	+3	—

Voice parameter changes for five emotions available for the Eloquent TTS system. Speech rate is words per minute (WPM). Average pitch (AP) in Hz. Loudness (G5) in dB.

SCREAM

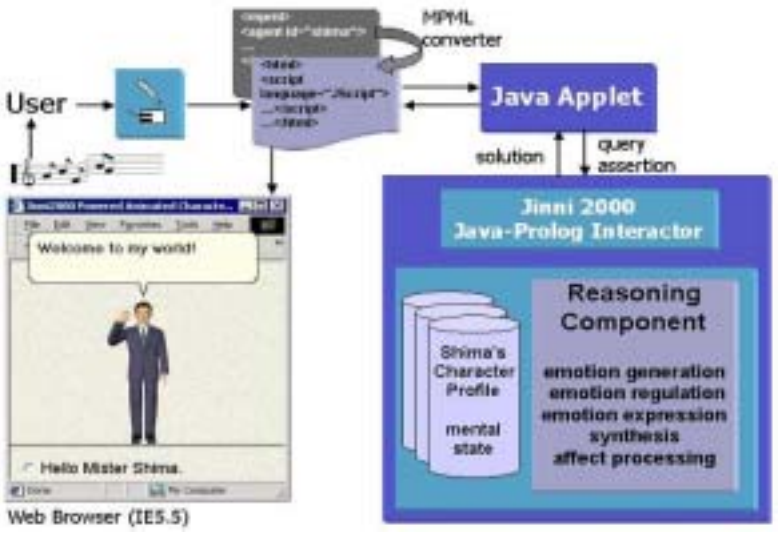
SCRipting Emotion-based Agent Minds



⌘ Emotions are derived from an agent's beliefs, goals, attitudes, and expressed according to personality trait (agreeableness, extroversion).

SCREAM -- implementation

SCRipting Emotion-based Agent Minds



Emotion Regulation in *SCREAM*

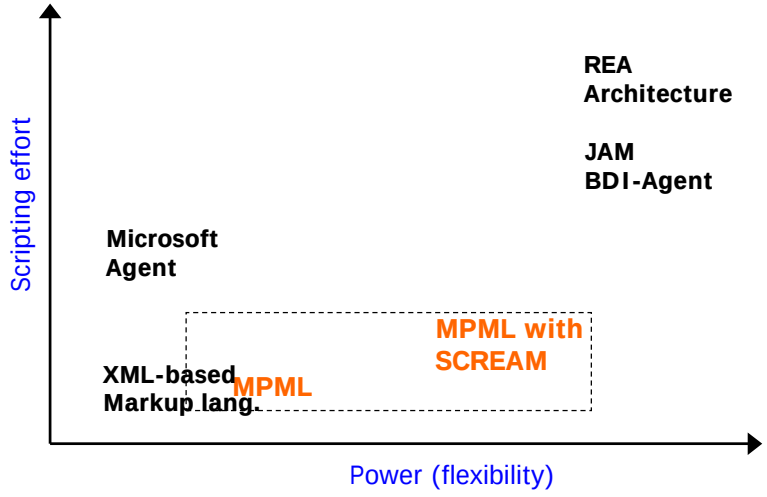
- **Ekman & Friesen's (69) "display rules"**
  - Expression of emotional state is governed by social and cultural norms, intensity of facial expression
- **Brown & Levinson (87) on linguistic style**
  - Assessment of seriousness of Face Threatening Acts (FTAs) considering agent's desire for autonomy and approval
  - Social variables: distance, power, imposition of speech acts
  - … avoiding disharmony in conversation (Moulin 98)
- **Poggi & Pelachaud (01) on reflexive agents**
  - Hamlet module: to display or not to display (emotions)
- **J.Gross (98) on emotion regulation in psychology**

  Emotion regulation refers to processes by which individuals influence which emotions they have, when they have them, and how they experience and express these emotions. (p.275)

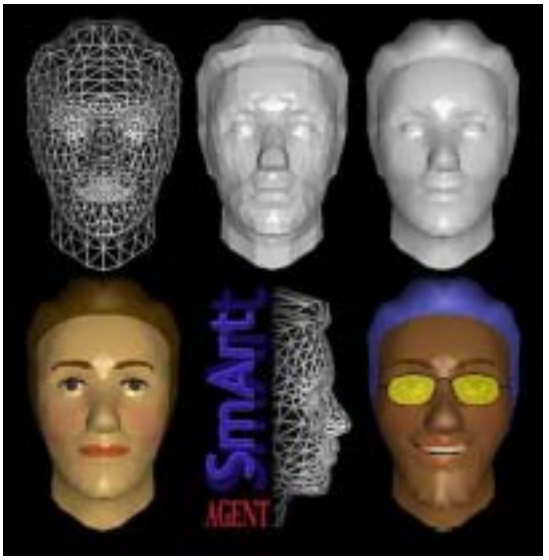
Interface between *MPML* and *SCREAM*

```
<!--MPML script illustrating interface with SCREAM -->
<mpml>
…
  <consult target="[…].jamesApplet.askResponseComAct('james,'al','5')">
    <test value="response25">
      <act agent="james" act="pleased"/>
      <speak agent="james">I am so happy to hear that.</speak>
    </test>
    <test value="response26">
      <act agent="james" act="decline"/>
      <speak agent="james">We can talk about that another time.</speak>
    </test>
    …
  </consult>
…
 </mpml>
```

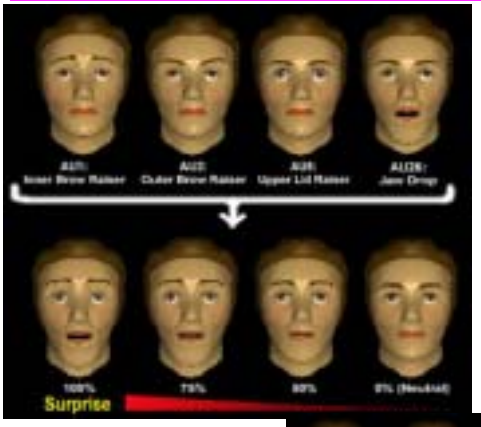
SCREAM technology as a scripting language



SmArt Agent



SmArt Agent's Faces
with Emotion Expressions

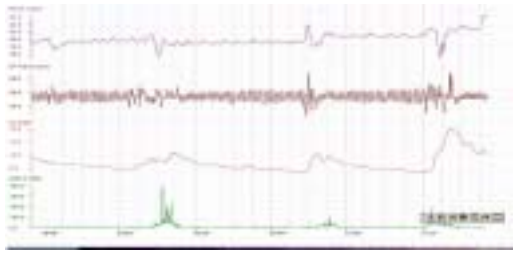


Application to 3D Chat with Emotional Expressions

3D Chat
= facial exp. recognition
+ agent + chat

Biophysical Emotion Sensors for Affective Interaction

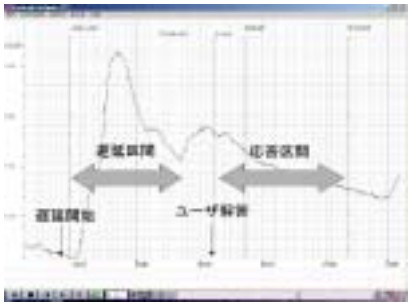
- **Skin-conductivity** (associated with **Arousal**)
- **Heart-pulse rate** (associated with **Valence**)
- **Others**
  - **Blood pressure, Temperature, Breath rate,**
  - **Electocardiogram(ECG), Brain waves(EEG), Electromyography(EMG)**



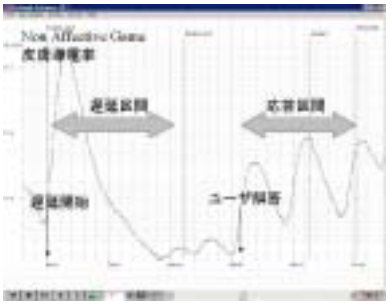
Original Biophysical Emotion Sensing Device with Bluetooth Interface

- We have developed our original emotion sensing device with the bluetooth wireless interface to detect:
  - SC (Skin Conduct)
  - HR (Heart rate)
- We tested it in the learning process, etc.

Effects appeared as Skin-conductivity

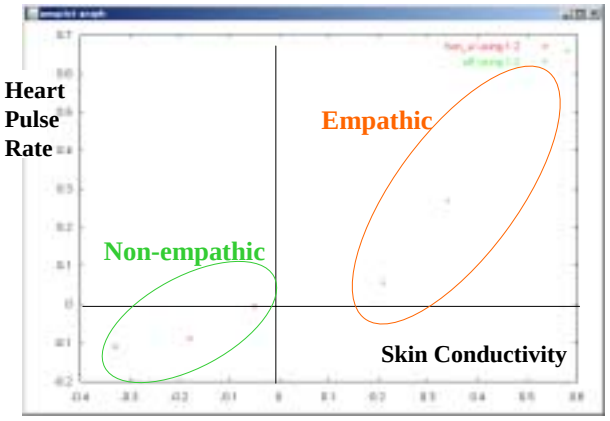


empathic interaction

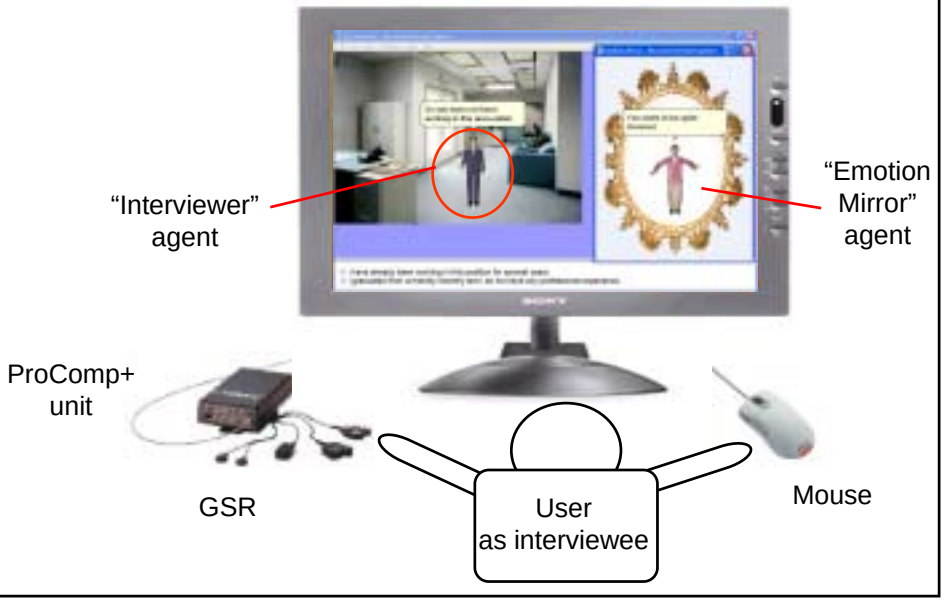


non- empathic interaction

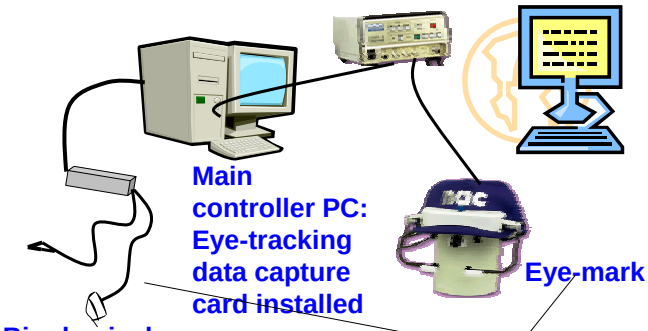
Persona Effect Observation in Empathic and Non- empathic Agents



Emotion Mirror in Virtual Job Interview



Eye-tracker in addition to biophysical sensors for affective interactions



Main controller PC: Eye-tracking data capture card installed

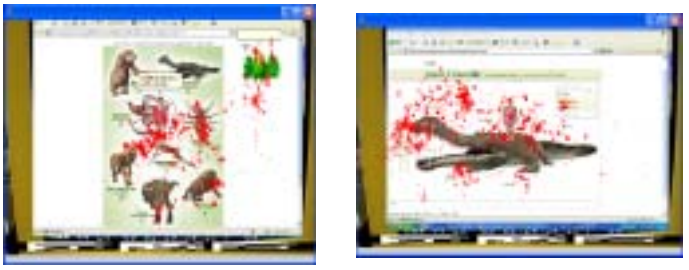
Eye-mark recorder

Biophysical sensors (skin conductance)

(students wear the both)

Analysis from Eye-tracking Data

Our approach vs. approach without agent reaction to eye-tracking



With

Without

English Conversation Training using *MPML* and Character Agents

Towards *MPML-mobile* version

- Small Display Area, Restricted Behaviors
- Menu Selection Inputs other than Voice or Text Inputs.
  ⇨ Redesign of MPML Tags.
- Contents generation through Mobile Java Appli.
  ⇨ MPML-mobile Converter to Mobile Java Appli.
- Small Memory (≦100KB) ⇨ Still big problem

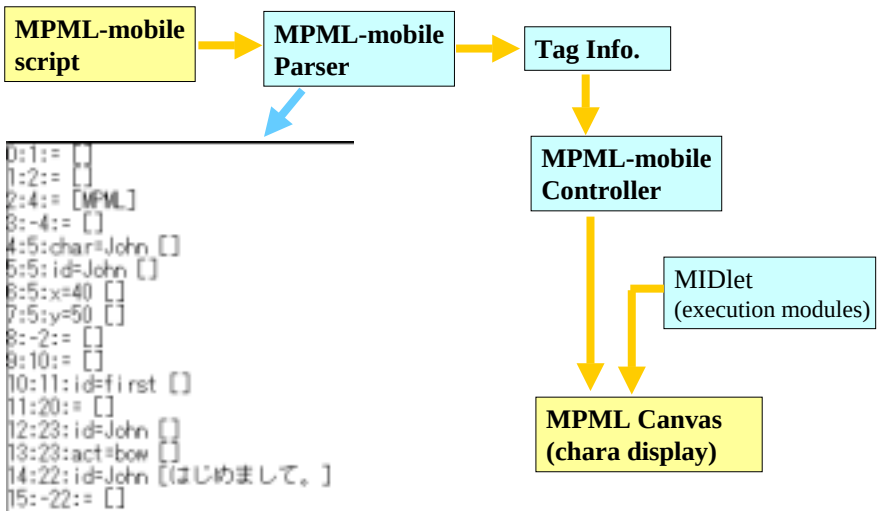
MPML-mobile0.5

```
<mpml>
  <head>
  <title>MPML</title>
  <agent char="John" id="John" x="40" y="50"/>
  </head>
  <body>
  <scene id="first">
    <seq>
      <play id="John" act="bow" />
      <speak id="John">はじめまして。</speak>
      <speak id="John">ボクの名前はJohnです。</speak>
      <par>
        <play id="John" act="bow" />
        <speak id="John">よろしく。</speak>
      </par>
      <pause/>
      <speak id="John">今までに撮った写真を紹介します。</speak>
    </seq>
```

3D Characters for Mobile Phones

⌘ **Based on Hi-Corp's Mascot Capsule Engine**
(a light-weight 3D modeler for mobile phones)

MPML-mobile ➔ J2ME on mobile phone



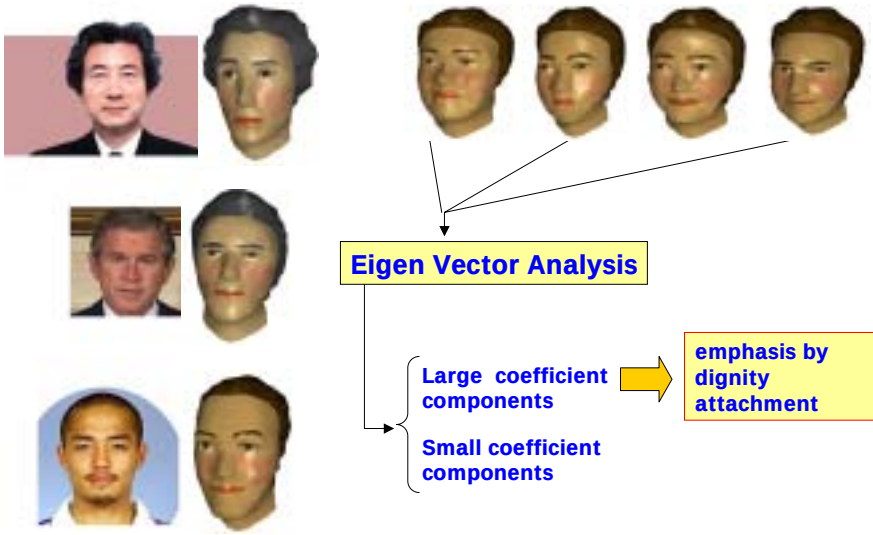
MPML mobile version

```
<?xml version="1.0" encoding="shift_jis"?>
<?xml:stylesheet type="text/xsl" href="mpml.xsl"?>
<mpml>
  <head>
    <title> Hello World! </title>
    <agent char="rockey" id="rockey" x="400" y="100"/>
  </head>
  <body>
    <par>
      <play id="rockey" act="Nod"/>
      <speak id="rockey"> Hello World!
          You're ready to proceed </speak>
    </par>
  </body>
</mpml>
```

MPML-mobile for *KDDI-au's* EZ-web, *DoCoMo's* i-mode, and *Vodafone*

in cooperation with Hottolink Inc.

Portrait Character Synthesis



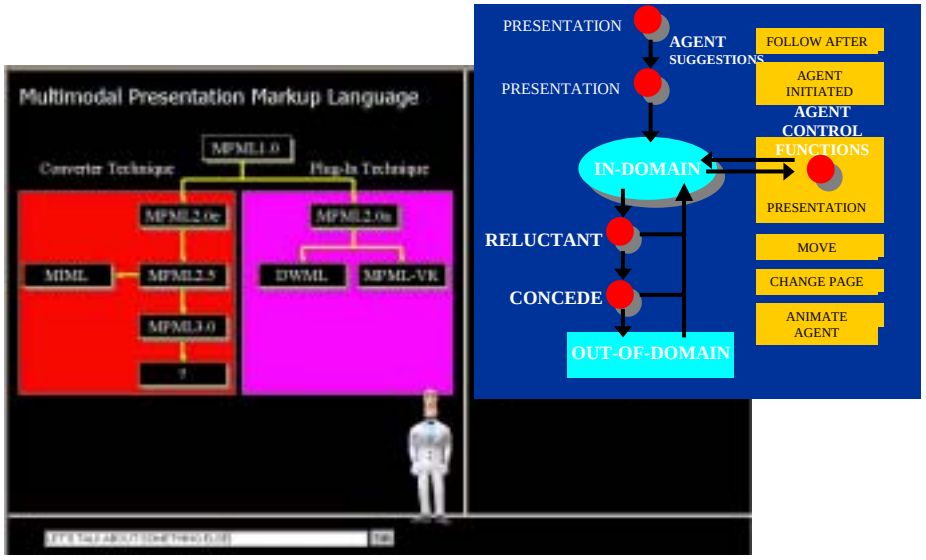
for Mobile Phone (i-mode)

- At present, the Cartoon-like Portrait SmArt Agents run only on D504i which has a graphic hardware.

Flexibility is not enough at present in Interactive Dialogue

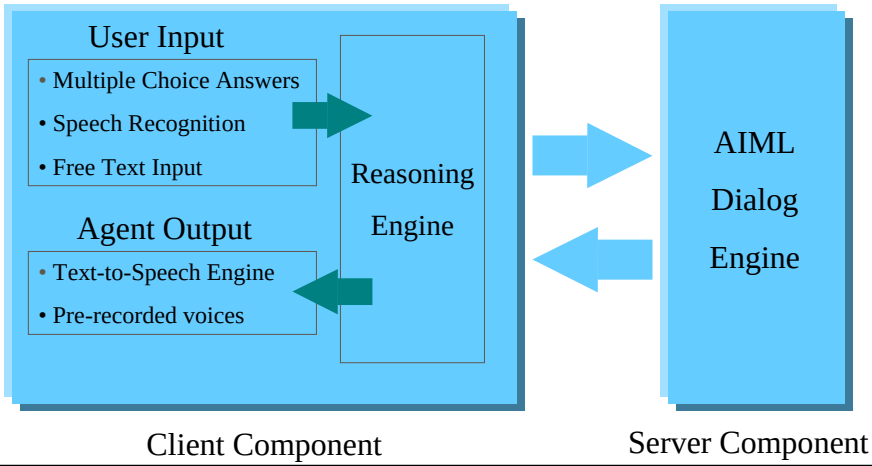
Technique used	Example Task	Task Complexity	**Dialogue Phenomena Handled**
Finite-state script	Long-distance calling	Least complex	**User answers questions**
Frame based	Getting train arrival/dept. info		**User asks questions, simple clarifications by system**
Set of contexts	Travel booking agent		**Shifts between predetermined topics**
Plan-based models	Kitchen design consultant		**Dynamically generated topics, collaborative negotiation sub-dialogues**
Agent-based models	Disaster relief management	Most complex	**Different modalities, planed/actual world**

Dialogue Control -- switching between In-domain and Out-of-domain (Chatterbot)



Enhancement of Conversational Flexibility through Chatbot technology

- **ALICE Chatbot**
  (by Richard Wallace, Winner of the 2000&2001 Loebner prizes)
- **AIML** (Artificial Intelligence Markup Language)



Auto Presentation with Web Intelligence Functions

- Understand the presentation topic from input query.
- Search the topic in Wikipedia, or Search by Google, Yahoo and AltaVista.
- Text segment summarization (extraction) , and associate with relevant outline.
- Generation of a scene-based *MPML* script with affective support.

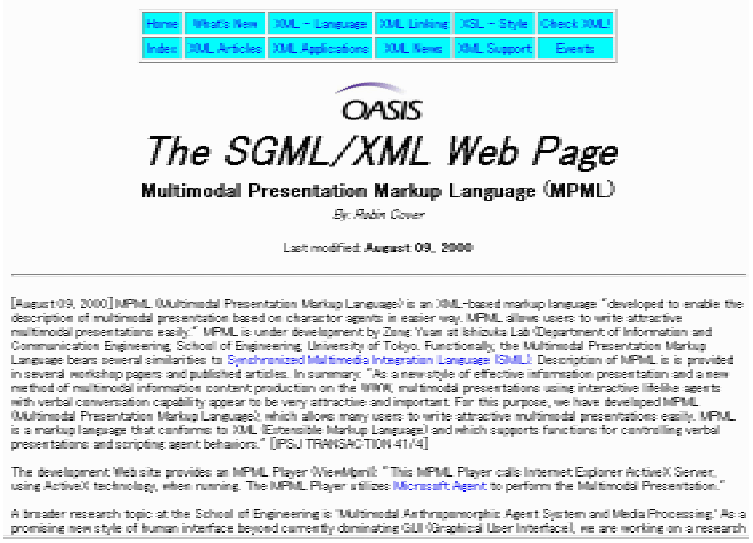
The topic is “Big Bang” here.

MPML basic tools are available at

http://www.miv.t.u-tokyo.ac.jp/MPML/

MPML's International Publicity

by *Robin Cover* at http://www.oasis-open.org/cover/mpml.html



Home | What's New | XML – Language | XML Linking | XSL – Style | Check XML!
Index | XML Articles | XML Applications | XML News | XML Support | Events

OASIS

The SGML/XML Web Page

Multimodal Presentation Markup Language (MPML)

By: Robin Cover

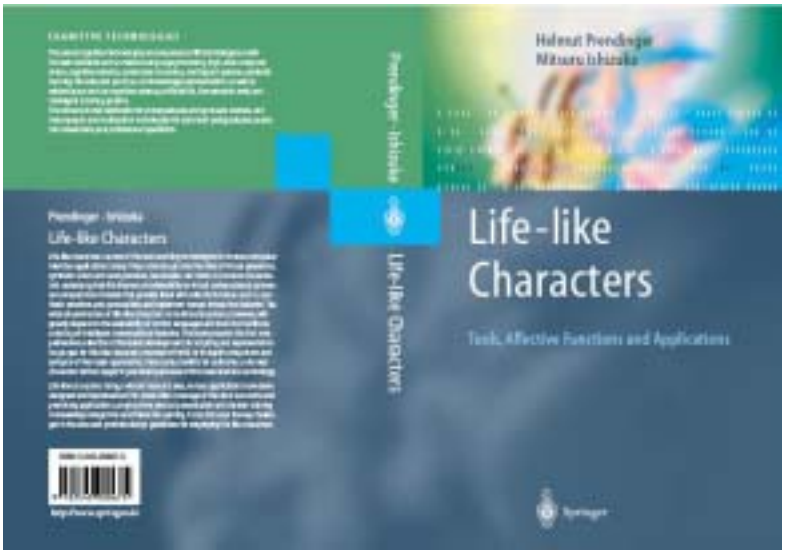
Last modified: August 09, 2000

[August 09, 2000] MPML (Multimodal Presentation Markup Language) is an XML-based markup language "developed to enable the description of multimodal presentation based on character agents in easier way. MPML allows users to write attractive multimodal presentations easily." MPML is under development by Zeng Yuxin at Ishizuka Lab (Department of Information and Communication Engineering, School of Engineering, University of Tokyo. Functionally, the Multimodal Presentation Markup Language bears several similarities to Synchronized Multimedia Integration Language (SMIL). Description of MPML is is provided in several workshop papers and published articles. In summary: "As a new style of effective information presentation and a new method of multimodal information content production on the WWW, multimodal presentations using interactive lifelike agents with verbal conversation capability appear to be very attractive and important. For this purpose, we have developed MPML (Multimodal Presentation Markup Language), which allows many users to write attractive multimodal presentations easily. MPML is a markup language that conforms to XML (Extensible Markup Language) and which supports functions for controlling verbal presentations and scripting agent behaviors." [IPSJ TRANSACTION 41(4)]

The development Website provides an MPML Player (WinMpml): "This MPML Player calls Internet Explorer ActiveX Server, using ActiveX technology, when running. The MPML Player utilizes Microsoft Agent to perform the Multimodal Presentation."

A broader research topic at the School of Engineering is "Multimodal Anthropomorphic Agent System and Media Processing." As a promising new style of human interface beyond currently dominating GUI (Graphical User Interface), we are working on a research

Our edited Book published from Springer in 2004



Summary of the Talk

- Background and Related Work
- Overview of MPML
- Various Versions of MPML
- Emotion Expressions (SCREAM)
- An Original Character Agent with Rich Expressions (SmArt)
- Conversational Flexibility
- MPML-VR (virtual reality)
- MPML-mobile
- MPML-HR (humanoid robot)
- Applications (web presentations, entertainments, language learning, etc.)

Current Issues

- **More Autonomy**
  - Extraction of Emotion from Texts ➔ Emotional Behavior Generation.
  - A Combination of a Chatbot for flexible conversation.
  - Behavior Plan Generation based on the Intention, Goal …. of an Agent.
  - Storytelling.
- **Affective Communication**
  - Emotion Sensors (face, voice, skin conductivity, …).
  - Modification of Output Sentences.
- **Multimodal Content Business**
  - Mobile Contents.
  - Multimodal Educational Contents.

Acknowledgments to the MPML members

Helmut Prendinger
Hiroshi Dohi
Santi Saeyor
Istvan Barakonyi
Sylvain Decamps
Kyoshi Mori
Arturo Nakasone
Mostafa Al Aasum
Masanori Mitsuzumi
Jun’ichiro Mori
Du Peng
Naoaki Okazaki
Yuan Zhong
Zhenglu Yang
Takayuki Tsutsui
Ma Chunling