

Proposal of a Script Language for Embodied Conversational Agents as Asynchronous Conversational Media

Toru TAKAHASHI^{†, ‡}

Hideaki TAKEDA^{*, †}

[†] Graduate School of Information Science, Nara Institute of Science and Technology,
8916-5 Takayama-cho, Ikoma-city, Nara 630-0101, JAPAN
+81 743 72 5265

[‡] ATR Media Information Science Laboratories,
2-2-2 Seika-cho, Soraku-gun, Kyoto 619-0288, JAPAN
+81 774 95 1486

* National Institute of Informatics,
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, JAPAN
+81 3 4212 2543

toru@atr.co.jp

takeda@nii.ac.jp

ABSTRACT

In this paper, we propose a script language for *Personified Media* (PM). PM is a kind of embodied conversational agent that can function as a personal conversational medium in asynchronous community systems. In our community system, participants have their personal PM to communicate with others. Each participant composes multimodal utterances towards his/her own PM in a script language and submits the utterances to the community system by turns. Because the script language of PM is not a language for agent system developers but for participants for online communities, the language should have a structure that users can compose utterances easily and naturally.

Categories and Subject Descriptors

H.4.3 [Information Systems Application]: Communications Applications; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces

General Terms

Design, Languages

Keywords

interface agents & conversational agents, agent language and environments, online community, asynchronous communication

1. INTRODUCTION

When communicating with others, people represent their intentions integrating their full-body modalities such as voice, gesture, pose, gaze, and motion. Additionally, people use shared context around them including information brought by the media. In online communities formed with information systems such as newsgroups and bulletin board system (BBS), however, people usually communicate only by text. There, people communicate without richly expressive faces and hands and only in words. People implicitly share web contents by including links to web pages; people cannot explicitly share the web pages while the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AAMAS'02, July 15-19, 2002, Bologna, Italy.
Copyright 2002 ACM 1-58113-480-0/02/0007...\$5.00.

communities are located in the “information space.”

Our goal is to make online communities more communicative and more social place, and we think *embodied conversational agents* (ECAs) [1] are one of the most probable solutions. In an asynchronous online community system like TelMeA [2], one or more ECAs are employed as personal conversational media. In this paper, we use the term *Personified Media*, or PM, for the ECAs that function as personal conversational media. Asynchronous conversation with PM enables people to communicate reflectively through multimodal behavior and to explicitly share web pages among them.

Most script languages used to describe the behaviors of autonomous interface agents are designed to describe a scenario of interaction logically and consistently. That is, they are languages for developers of agent systems. On the other hand, script languages for PM are languages for users to communicate. Therefore, the languages for PM should be designed so that users can describe utterances naturally and adequately.

In this paper, we propose a new script language named PMScript. One PMScript description represents one utterance, and conversations are built through asynchronously exchanging the PMScript descriptions instead of the texts in the case of newsgroups and BBS. PMScript supports and enhances natural social interaction in a community, by enabling its users to create multimodal utterances easily.

2. PROPOSAL FOR A NEW SCRIPT LANGUAGE FOR PM

PMScript is currently provided in XML syntax. Figure 1 shows the structure tree of PMScript. As the figure shows, the content of PMScript consists of two elements, i.e., CAST element and SCRIPT element.

2.1 CAST element

The CAST element presents a list of PM and multimedia contents that are involved in the Script element. PM of the list are presented in the PM element and multimedia contents of the list are presented in the CONTENT elements.

The CAST element has only one attribute, that is, the ID attribute. The ID attribute is an identifier of elements. Most elements in PMScript have this attribute, and it simply shows the identifier, so we will omit the explanation of the ID attribute hereafter.

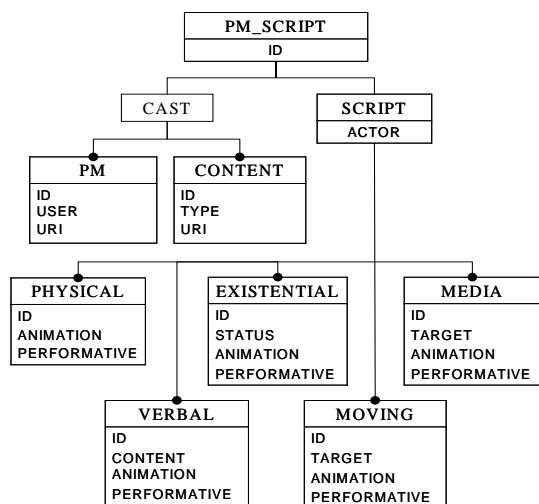


Figure 1. Structure tree of PMScript

2.2 PM element and CONTENT element

The PM element presents PM of a participant in this script, and the CONTENT element presents multimodal content as well. The PM element and CONTENT element each have three attributes: ID, USER, and URI attributes for the PM element, and ID, TYPE, and URI attributes for the CONTENT element. The USER attribute of the PM element presents the user of the PM. The TYPE element of the CONTENT element presents the MIME content type of this content. The URI attribute of both elements presents the URI (Uniform Resource Identifier) of this PM/content.

2.3 SCRIPT Element

The SCRIPT element has only one attribute, The ACTOR attribute. The utterer PM of the script is designated by this ACTOR attribute in the SCRIPT element. The SCRIPT element presents the context of the utterance as a combination of its sub-elements, i.e., VERBAL, PHYSICAL, MOVING, EXISTENTIAL, and MEDIA elements. Each element presents representations by PM and they are classified by modalities. Each element has ID ANIMATION, and PERFORMATIVE attributes. The ID attribute is an identifier of each element. The ANIMATION attribute designates PM's animation including, for example, lip-sync animation when PM speaks verbally and stepping animation when PM moves around. The PERFORMATIVE attribute indicates intention of performing the representation. The purpose of this attribute is to clarify the intentions of speakers when they compose the script. For instance, we suppose users select their behavior not from the representation of animations but from the PERFORMATIVE, that is, the intention of this behavior. The same behavior often has different meanings in various contexts.

We explain details of each element below.

2.4 PHYSICAL Element

The PHYSICAL element presents a representation of a physical modality by animation, such as gestures, facial expressions, and poses. The PHYSICAL element has ID, ANIMATION, and PERFORMATIVE attributes. The ANIMATION attribute indicates the entity of the animation, which shows a sequence of physical movement of PM. On the other hand, the PERFORMATIVE attribute presents the user's intention which

the user intend to express by performing this physical representation. For instance, when a "smile" animation is selected in a utterance, the PERFORMATIVE attribute indicates that this "smile" could specify either an acceptance or a sarcasm.

2.5 VERBAL Element

The VERBAL element presents linguistic representation such as balloon text and synthesized voice via PM. In addition to ID, ANIMATION, and PERFORMATIVE, the VERBAL element has CONTENT attribute. The CONTENT attribute presents the linguistic sentences for this representation.

2.6 MOVING Element

The MOVING element presents representation through the movement of the PM on the display. The MOVING element has TARGET attribute other than ID, ANIMATION, and PERFORMATIVE. The TARGET attribute presents the movement destination point of the PM, which is specified in terms of either an ID attribute of a target PM or a target CONTENT element, or the coordinate value of a target position on the display.

2.7 EXISTENTIAL Element

The EXISTENTIAL element presents a change of a state of a PM, such as size of PM, appearance of PM, or the state of whether the PM is displayed or not. The EXISTENTIAL element, for example, may make the PM bigger, reclothe the PM, or hide the PM. In addition to ID, ANIMATION, and PERFORMATIVE, the EXISTENTIAL element has a STATE attribute. The STATE attribute presents the state of PM after this element is performed.

2.8 MEDIA Element

The MEDIA element presents representation with multimedia contents. In addition to ID, ANIMATION, and PERFORMATIVE, the MEDIA element has TARGET attribute. The TARGET attribute presents the target of a multimedia content to operate such as letter string, image, movies, web pages, and so on. This value of the TARGET attribute is designated as an ID attribute of the CONTENT element.

3. CONCLUSION

In this paper, we propose a script language for ECA that can function as the personal conversational media on asynchronous community systems. The benefits of PMScript are that it (1) provides a standard protocol that enables different PM-based community systems, and (2) provides a good structure of utterances in order to reuse them. Future works around PMScript is practice and evaluation. We are currently working to adapt our new community system (cf. [2]) using PM to be able to understand and process PMScript, and we are going to propose to the agent research community that the specification of PMScript be used as the standard in order to stimulate further discussion.

4. ACKNOWLEDGEMENT

The authors wish to thank Dr. Yasuhiro Katagiri of ATR Media Information Science Labs for making constructive comments for improving this paper. This research was supported in part by the Telecommunications Advancement Organization.

5. REFERENCES

- [1] Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. (eds.), *Embodied Conversational Agents*, Cambridge; MIT Press, 2000.
- [2] Takahashi, T. and Takeda, H., *TelMeA: An Asynchronous Community System with Avatar-like Agents*, Proc. INTERACT2001, pp. 480-487, 2001.