Aspects of Interface Agents: Avatar, Assistant and Actor

Kenji Mase
ATR Media Integration & Communications Research Labs, Kyoto, Japan
mase@mic.atr.co.jp

Abstract

This paper introduces the interface agent research project being carryied out at ATR Media Integration & Communications Research Labs. We are interested in virtual interface agents to support human creative activities by mediating between human users and computer cyberspace. In this paper, we categorize interface agents into three types; *avatars, assistants and actors* based on their functionality and discuss a design framework of agents and related applications. We present the current status and objectives for the following topics; design of an asynchronous-hierarchical agent architecture, a character locomotion design tool, applications in a virtualized museum and a group discussion environment in which virtual agents would inhabit and act.

1 Introduction

A "virtual interface agent," or interface agent, is defined in this paper as an autonomous agent which mediates between a human user and computer cyberspace. An interface agent differs from an ordinary interface since it is expected to change its behaviors and actions autonomously according to the human user's behaviors and the situations as the interaction progresses. At the same time, we expect the agent to maintain a consistent, defined personality or *whole character* even though it displays various behaviors in different situations depending on the context. This is particularly important when the agent is designed as a personal interface agent. We speculate that, in the future, an agent will accompany the user all day performing a role that a human butler or a secretary would at home or the office. Such a personal interface agent should have a solid personality as the foundation of its character but also have the flexibility to play different roles for the sake of its agentee. Further, as the agent adapts to the agentee (and vice versa) we foresee the relationship between them will become more intimate over time.

We are designing a multi-agent architecture which is suitable to develop such interface agents for various application systems. The architecture is also unified to control animated agent characters[Bruderlin *et al.*, 1997]. Without any practical tasks, an interface agent would remain a toy to play or a friend chat to. We are pursuing two applications; a virtualized museum[Mase *et al.*, 1996] and a creative discussion support environment system[Nishimoto *et al.*, 1997][Sumi *et al.*, 1997]. Agents will appear in both systems and play active roles.

As we refer to a person differently by personal pronouns, such as *I, you and they*, in a segment of discourse, we can refer to an interface agent from different perspectives depending upon their relationship to us (and our task) during discourse. Three categories of agents we consider are avatars, assistants (or secretary agents) and computer actors. Unlike the personal pronouns however, this categorization of an agent, which we call *agentality*, identifies an agent based upon its function.

The categorization helps us in designing interface agents. In a virtual museum application, for instance, a user may want the agent to play three kinds of roles during the visit. First, an avatar in the virtual world is helpful to extend his/her reality and grasp the situation in the virtual world. Second, a virtual assistant (guide) agent may answer questions about the visit and bibliographical inquiries. Third, a virtual actor can entertain the visitor by showing a play. An avatar may become an actor when such a show is open for participation to the virtual visitors. The largest difference among the types of agents is the level of control accessible by user. The user can make an agent be an avatar, assistant or actor by simply changing the level of control he/she exercises. But still he/she expects that the avatar and the actor will maintain an identical personality. In this way, the agent provides the user with a consistent interaction manner while allowing him/her to utilize different agentality. It is still an open issue how to display the change of agentality. One possibility has the figure metamorphosis when agentality changes; however, this approach may cause difficulties with the cognitive process of keeping track of identical personality. Another possibility is that users may be aware of the change in status by using the context of the interactions themselves. This is similar to the way people have the ability to change their mode of interaction with other people depending upon

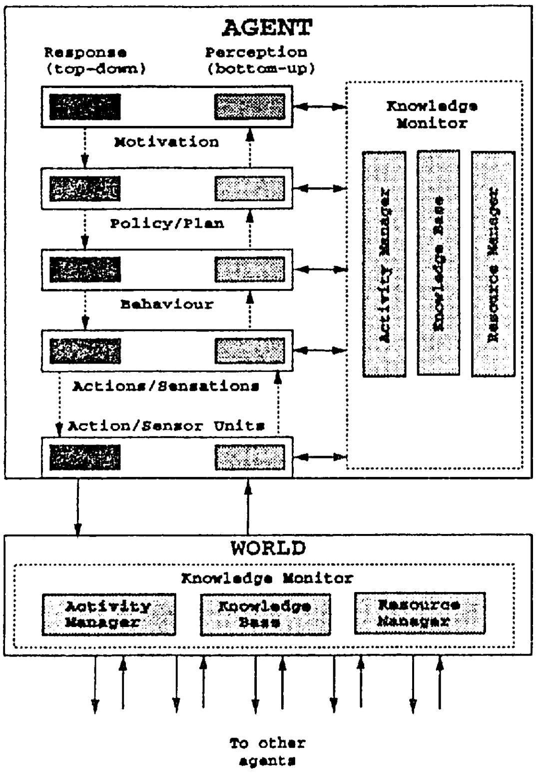

Figure 1: A-HA: Asynchronous-Hierarchical Agent architecture

context.

The multi-agent architecture we are designing has a multi-layered structure based on a hierarchical model [Fischer *et al.*, 1995] which allows for a user to intervene at any level of the hierarchy. It is named Asynchronous-Hierarchical Agent (A-HA) architecture [Fels *et al.*, 1996]. The hierarchy is basically composed of 5 levels: motivation/needs, plan/policy, behavior, actions/sensations and action/sensor units. We have the DOF (degree of freedom) layer, strictly speaking, under the action unit layer to control agent figures geometrically. This architecture fits with our desire to design different agentalities in a uniform framework.

The visual embodiment of an agent is also an important issue to provide visual feedback of agent's action. The interactive real-time locomotion design interface [Bruderlin and Mase, 1996] is currently an animation tool rather than a visual representation control system of an animated agent. However, it is designed based on the same philosophy of the A-HA architecture and responds to interactive control from users at the behavior and the action layers.

In the following sections, we discuss our on-going interface agent research projects which use our agent philosophy and are candidates for using agents built around the A-HA architecture. In section 2 we expand on how the A-HA architecture works. In section 3 we present the locomotive animation mechanism of visual agents which is available on the A-HA platform. In sections 4.1 and 4.2, a few applications are discussed; a virtualized museum, in which a personal agent will guide the visit of a museum, and a creative discussion support environment system, in which a discussion support agent will attend an online meeting as a virtual participant.

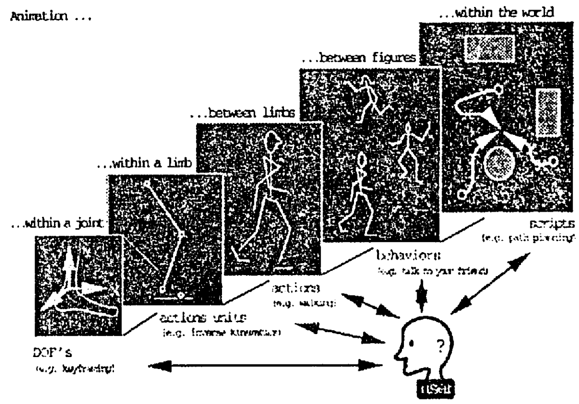

Figure 2: A Hierarchical animation control interface

2 A-HA Architecture

Two fundamental components are in the Asynchronous Hierarchical Agent (A-HA) architecture as shown in Figure 1. First, the multilevel hierarchical structure of the system is comprised of motivation/needs, policy/planning, behaviors, action/sensation and action/sensor units. Each layer contains the responsive and perceptual functions, which interpret input signals from the adjacent layer. Our approach resembles the multi-layered approach found in the InteRRaP system [Müller *et al.*, 1994]. In our system, we have a structure called the knowledge monitor which contains a knowledge base that is structured so that all control layers have access to the information contained; providing access to knowledge that applies to all levels of control abstraction. Knowledge specific to a particular layer though is contained within that layer. Our agent architecture allows the user to control the agent at different levels of abstraction. The various levels of abstraction correspond to different agentalities allowing uniform design of avatars, assistants and actors. Agents interact with each other through the world, where the shared knowledge monitor is managing common activities and a common knowledge base. The knowledge monitor of the world acts like the blackboard of our multi-agent systems.

An example of designing each agentality follows. The avatar requires frequent control at the action layer by the user. It does not have an explicit objective or sense of intelligence but just acts like a marionette. The agent, however, adapts the action unit so that dynamic actions looks natural. The assistant agent expect explicit inputs of goals and plans from the user. It watches the behaviors of users as action examples or uses sensed situations about the user and the environment. The basic behaviors and layers beneath are pre-coded. The actor agent will follow the pre-coded script most of time looking at

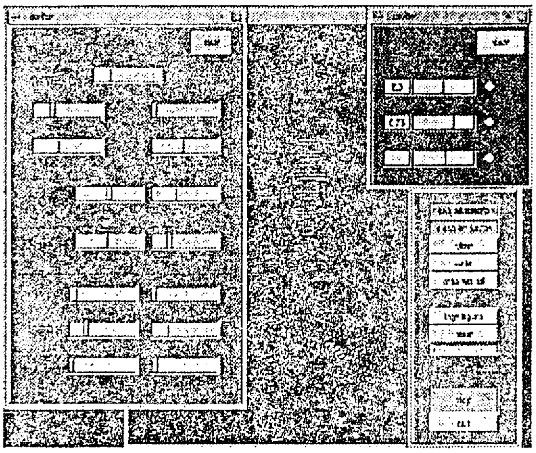

Figure 3: Gaitor: an animation design tool for locomotion

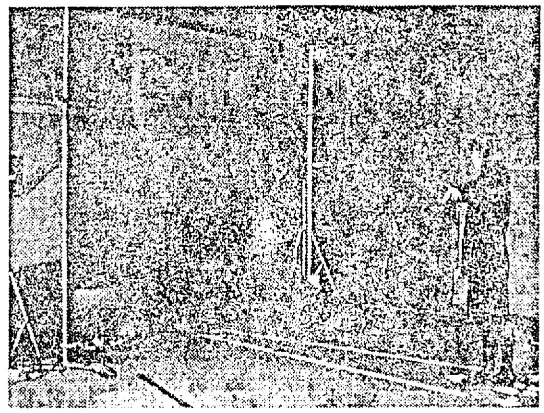

Figure 4: A scene from VisTA-Walk system: gesture-based walk-though and information access in a virtual ancient village. TV camera is set on the top of screen.

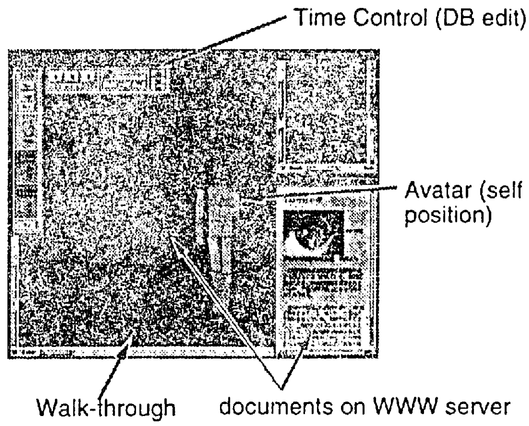

Figure 5: A screen shot of VisTA: an avatar character appears as a position reference in a virtual ancient village walk-through.

the human user and other agents and responds properly within the given freedom to proceed in the script.

The A-HA architecture is being tested in the implementation of agent card players in a simple card game called Agent 99 [Fels *et al.*, 1996] and a Malefitz board game. After the feasibility of the architecture is confirmed, we will start to implement other real applications described below.

3 Animation Control of Visual Agents

Realistic and/or believable motion is important to give an agent an intelligent appearance. Unfortunately, it is very tedious and requires great skill to produce realistic and believable motion by traditional key-framing. As the actions of an agent become more complex, it is not practical to design all agent actions using the labor of a human animator nor by motion capturing. An algorithm-driven animation process is strongly desired.

By incorporating knowledge about how real humans move into the control algorithms, certain motions can be animated more convincingly, autonomously and efficiently. Interestingly, the animation control interface can be layered hierarchically similar to A-HA architecture (see Figure 2). We have developed an interactive, real-time animation system of human locomotion (walking and running) which has two level access control corresponding to the behavior layer and the action layer. [Bruderlin and Mase, 1996]. Figure 3 is an interface display of the system. In the locomotive agent, walking speed, step length, etc. are the behavior layer parameters, and other attribute parameters such as torso bend and pelvic rotations are the action layer parameters.

This animation system has been developed as an independent system and is being incorporated in an A-HA architecture to make visual agents autonomously movable in the cyberspace of various applications.

4 Applications

In the next two subsections, We introduce two application systems in which autonomous agents will inhabit and interact with human users.

4.1 Meta-Museum

First, we have proposed a Meta-Museum concept [Mase *et al.*, 1996]. The Meta-Museum is a interdisciplinary museum incorporating augmented reality, artificial intelligence and interface agent technology to facilitate the communication between visitors and people behind the exhibits. The objective of Meta-Museum is to provide an interactive, exciting, entertaining and educational experience for visitors. The Meta-Museum offers rich and effective interaction with museum's archives and the people behind them mediated by interface agents.

Figure 4 shows the first prototype subsystem of Meta-Museum named *VisTA-Walk* (VisTA: Visualization Tool for Archaeological data), which is a virtual walk-through system with vision-based input devices. We have mod-

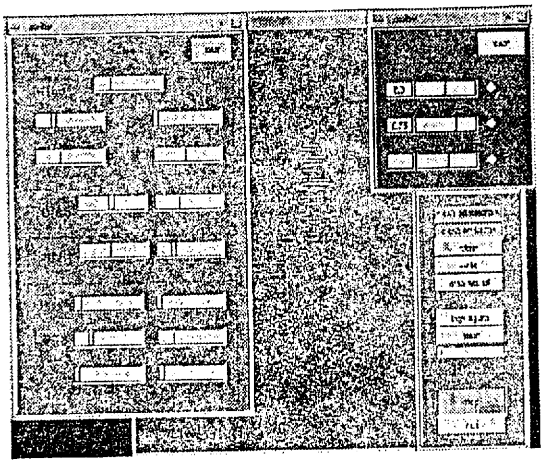

Figure 3: Gaitor: an animation design tool for locomotion

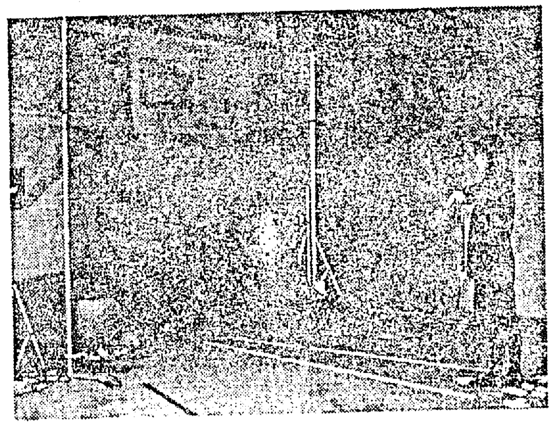

Figure 4: A scene from VisTA-Walk system: gesture-based walk-though and information access in a virtual ancient village. TV camera is set on the top of screen.

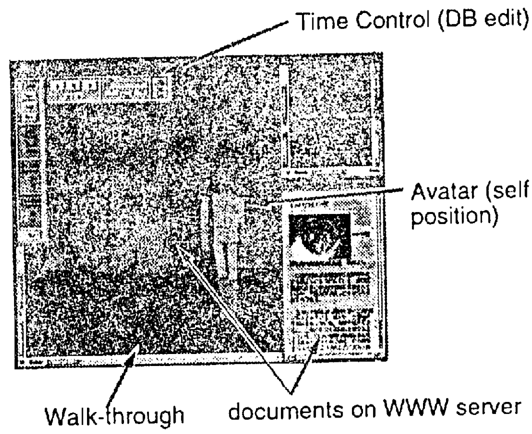

Figure 5: A screen shot of VisTA: an avatar character appears as a position reference in a virtual ancient village walk-through.

the human user and other agents and responds properly within the given freedom to proceed in the script.

The A-HA architecture is being tested in the implementation of agent card players in a simple card game called Agent 99 [Fels *et al.*, 1996] and a Malefitz board game. After the feasibility of the architecture is confirmed, we will start to implement other real applications described below.

3 Animation Control of Visual Agents

Realistic and/or believable motion is important to give an agent an intelligent appearance. Unfortunately, it is very tedious and requires great skill to produce realistic and believable motion by traditional key-framing. As the actions of an agent become more complex, it is not practical to design all agent actions using the labor of a human animator nor by motion capturing. An algorithm-driven animation process is strongly desired.

By incorporating knowledge about how real humans move into the control algorithms, certain motions can be animated more convincingly, autonomously and efficiently. Interestingly, the animation control interface can be layered hierarchically similar to A-HA architecture (see Figure 2). We have developed an interactive, real-time animation system of human locomotion (walking and running) which has two level access control corresponding to the behavior layer and the action layer. [Bruderlin and Mase, 1996]. Figure 3 is an interface display of the system. In the locomotive agent, walking speed, step length, etc. are the behavior layer parameters, and other attribute parameters such as torso bend and pelvic rotations are the action layer parameters.

This animation system has been developed as an independent system and is being incorporated in an A-HA architecture to make visual agents autonomously movable in the cyberspace of various applications.

4 Applications

In the next two subsections, We introduce two application systems in which autonomous agents will inhabit and interact with human users.

4.1 Meta-Museum

First, we have proposed a Meta-Museum concept [Mase *et al.*, 1996]. The Meta-Museum is a interdisciplinary museum incorporating augmented reality, artificial intelligence and interface agent technology to facilitate the communication between visitors and people behind the exhibits. The objective of Meta-Museum is to provide an interactive, exciting, entertaining and educational experience for visitors. The Meta-Museum offers rich and effective interaction with museum's archives and the people behind them mediated by interface agents.

Figure 4 shows the first prototype subsystem of Meta-Museum named *VisTA-Walk* (VisTA: Visualization Tool for Archaeological data), which is a virtual walk-through system with vision-based input devices. We have modeled an ancient Japanese village where visitors can experience archaeological hypothesis visually in VisTA-Walk.

Interface agents will support visitors as personal agents or exhibit agents. The personal agent has three agentalities and changes the major portion of its interaction to the visitor as she/he proceeds through the exhibition. For example, an assistant agent will appear to guide the walking course through the museum and provide information of interest to the visitor. An actor agent will show how the artifacts were used; visually traveling back to the original era and location.

Such a computer actor agent will play its role together with other computer and/or human actors based on a given script. In order to make the computer actor intelligent-looking, the timing of interaction is important. Timing is especially difficult to control when playing with a human actor because the agent's sensory devices do not know how long a human action takes during role-playing since the activity manager does not know a priori how long an actor takes play a part. We have created a prototype role-playing system *"SingSong"* at ATR, which is developed using a scripting method named Interval Script [Pinhanez *et al.*, 1997] suitable for such computer actor environment.[Pinhanez and Bobick, 1996]

An avatar agent is already imported into the system to provide the visitor's position in the virtual village. There are three windows on screen; a walk-through view, a bird's-eye view and a document browser (i.e. WWW browser). The WWW browser is hooked to the objects (e.g. houses, ware-houses) in the walk-through view to provide the prepared information as requested by user's selection by the gesture pointing or mouse-clicking.

The close-up of the VisTA screen is shown in Figure 5, where the additional menus are presented for the use of archaeological researchers and exhibit designers. A character figure can be seen in a view, which is the avatar agent. The picture is taken by specially setting the virtual camera position in front of the avatar to give the idea how the interface agent inhabit in the virtual world. Usually, the position of virtual camera is set at the point of avatar's eye position or in the sky for the bird's-eye view.

The immersive presentation shown in Figure 4 is one way of the exhibit of Meta-Museum. We have a plan to introduce a see-through HMD (head mounted display) device for personalized presentation which fits to display personal interface agents.

4.2 Creative Conversation Support

We are developing a system called AIDE (Augmented Informative Discussion Environment) which facilitates creative conversations between people engaged in collaborative work. AIDE is an online client-server type chat system designed with conversation spaces shared by the users [Nishimoto *et al.*, 1997](see Figure 6). This system can be both centralized/distributed and synchronous/asynchronous. AIDE is characterized by the following three subsystems: *Discussion Viewer, Conversationalist* and *Personal Desktop*.

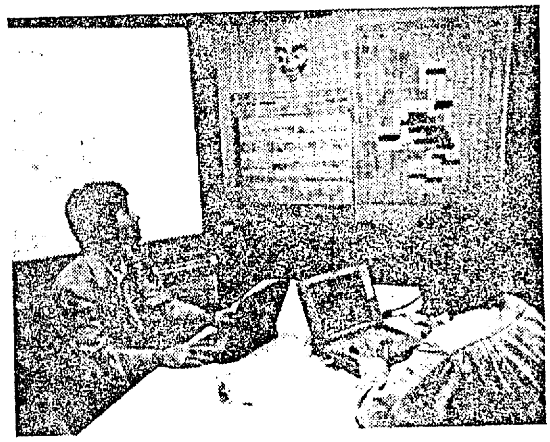

Figure 6: AIDE: a conversationalist agent in a group discussion supporting environment

The *Conversationalist* is being designed as an assistant agent who joins human conversations and provides additional and useful information. The agent provision is triggered manually by a user's request or automatically when the assistant determines appropriate by watching the conversation space. The agent acts as an information retrieval system which searches for the moderately relevant or the very relevant information. The choice depends on the situation of conversation. The agent is implemented mostly as a software agent and visual display is not necessarily needed in the current version. We speculate that this will change as the complexity of task and the amount of data increases. The Japanese Contemporary Dictionary is being used as the text data base of the Conversationalist, thus it only has a character of generalist currently.

The Discussion Viewer shows discussion spaces that visualize the structures of conversations. These spaces are information spaces shared among all participants in the conversations. They automatically visualize the relationships between utterances and their keywords using common keywords as the clue of measuring close relationships. The Personal Desktop is an viewer in which users can personalize and crystallize the shared information by duplicating and modifying the discussion spaces there.

5 Summary and Future Directions

In this paper the project on interface agent research at ATR-MIC has been presented. We are currently implementing the animated agent using the A-HA architecture for the applications outlined. We are gradually getting results on each sub-topic as described above. It is our belief that the vertically layered A-HA architecture and the agentality aspect of our agents will provide a convenient and clear mechanism for users and designers to interact with agents. The display and control of changing agentality has yet to be developed. We expect the structure of creative thought activity and changing perspectives in agent mediated communication systems can be described together with agentality.

This paper introduced the fundamental sub-projects of our interface agent research, i.e. A-HA architecture and agent locomotion algorithm. We plan to apply these results to application systems such as Meta-Museum and the AIDE system.

Acknowledgments

The author would like thank to Ryohei Nakatsu, for his support and encouragements in this research. He also would like to thank his colleagues, Armin Bruderlin, Silvio Esser, Tameyuki Etani, Sidney Fels, Rieko Kadobayashi, Kazushi Nishimoto, and Yasuyuki Sumi, for their substantial research contributions to the work mentioned here. Special thanks also to Peter Stucki and Sidney Fels for their comments to improve this article. The P-finder computer vision system used in the VisTA-Walk for gesture recognition is provided by the Perceptual Computing Section, Media Laboratory, MIT.

References

[Bruderlin and Mase, 1996] Armin Bruderlin and Kenji Mase. Interactive, real-time animation of human locomotion. In *Lifelike Computer Characters'96*, pages 73–74, Snowbird, October 1996.

[Bruderlin *et al.*, 1997] Armin Bruderlin, Sidney Fels, Silvio Esser, and Kenji Mase. Hierarchical agent interface for animation. In *IJCAI-97 Workshop on Animated Interface Agents*, Aug. 1997.

[Fels *et al.*, 1996] Sidney Fels, Kenji Mase, Armin Brudelin, Tameyuki Etani, and Silvio Esser. Agent 99: Implementing a simple card game using agents. In *The 2nd Intelligent Information Media Symposium*, pages 47–54. IEICE of Japan, Dec. 1996.

[Fischer *et al.*, 1995] K. Fischer, J. P. Müller, and M. Pischel. A pragmatic BDI architecture. In M. Woolridge, J.P. Müller, and M. Tambe, editors, *Intelligent Agents II*, pages 203–218. Springer, 1995.

[Mase *et al.*, 1996] Kenji Mase, Rieko Kadobayashi, and Ryohei Nakatsu. Meta-museum: A supportive augumented reality environment for knowledge sharing. In *Intn'l Conf on Virtual Systems and Multimedia'96*, pages 107–110, Gifu, Sept. 1996.

[Müller *et al.*, 1994] J. P. Müller, M. Pischel, and M. Thiel. Modelling reactive behaviour in vertically layered agent architectures. In M. Woolridge and N. R. Jennings, editors, *Intelligent Agents*, pages 261–276. Springer-Verlag, 1994.

[Nishimoto *et al.*, 1997] Kazushi Nishimoto, Yasuyuki Sumi, and Kenji Mase. Enhacement of creative aspects of a daily conversation with a topic development agent. In *Asian'96 post conference workshop on Coordination Technology for Collaborative Applications, Lecture Notes Series*. Springer-Verlag, Feb. 1997.

[Pinhanez and Bobick, 1996] Claudio S. Pinhanez and Aaron F. Bobick. Computer theater: Stage for action understanding. *AAAI'96 workshop*, Aug. 1996.

[Pinhanez *et al.*, 1997] Claudio S. Pinhanez, Kenji Mase, and Aaron Bobick. Interval Scripts: a Design Pradigm for Story-Based Interactive Systems. In *CHI97 Conference Proc.*, pages 287–294, Atlanta, GA, Mar. 1997.

[Sumi *et al.*, 1997] Yasuyuki Sumi, Kazushi Nishimoto, and Kenji Mase. Personalizing shared information in creative conversations. In *IJCAI-97 poster*, Nagoya, 1997.