



# **Interactive Movie System with Multi-person Participation and Anytime Interaction Capabilities**

Ryohei Nakatsu, Naoko Tosa, and Takeshi Ochi  
ATR Media Integration and Communications Research Laboratories  
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan  
Tel; +81-774-95-1400  
E-mail: nakatsu@mic.atr.co.jp

## **1. ABSTRACT**

**Interactive movies, in which interaction capabilities are introduced into movies, is considered to be a new type of media that integrates various media, including telecommunications, movies, and video games. In interactive movies, people enter cyberspace and enjoy the development of a story there by interacting with the characters in the story. In this paper, we first explain the concept of interactive movies using examples of movies developed on a prototype system, then describe techniques for improving the interactivity.**

**The current system incorporates two significant improvements for multimedia interactivity: the introduction of interaction at any time and two-person participation through the use of network communications. The software and hardware configurations of the system are briefly summarized. The paper concludes with an example of an interactive**

**story installed in this system and briefly describes the interaction between the participants and the system.**

### **1.1 Keywords**

Interactive movies, interactions, multi-story, anytime interaction, speech recognition, gesture recognition, emotion recognition,

## **2. INTRODUCTION**

Ever since the Lumiere brothers created Cinematography at the end of the 19th century[1], motion pictures have undergone various advances in both technology and content. Today, motion pictures, or movies, have established themselves as a composite art form in a wide domain that extends from fine arts to entertainment. Movies have the power to draw viewers into a virtual world where they can actually "experience" the development of a story. Based on the storytelling power that is embedded in novels and other forms of literature, movies add visual images and sound to create a virtual world (cyberspace).

The use of interaction technology, on the other hand, gives movies much greater inherent possibility than the current forms of movies. Conventional movies present predetermined scenes and story settings unilaterally, so audiences take no part in them and make no choices in story development. As a further step, the use of interaction technology makes it possible for the viewer to "become" the main character in a movie and enjoy a first-hand experiences. We believe that this approach would allow producers to explore the possibilities of a new class of movies.

Based on this viewpoint, we have been conducting research on interactive movie production by applying interaction technology to conventional movie making techniques. As an ini-

Permission to make digital/hard copies of all or part of this material for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copyright is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires specific permission and/or fee

©1998 ACM 1-58113-162-3/98/09 \$US 5.00

tial step in creating a new type of movie, we have produced a prototype system[4]. Based on this system, we are currently developing a second prototype system with many improvements. This paper briefly describes the first prototype system and outlines its problem areas and required improvements. The paper also introduces the configuration of the second prototype system, which is now under development by incorporating the described improvements.

### 3. OUTLINE OF INTERACTIVE MOVIES

#### 3.1 Concept

Compared with existing media, interactive movies can be regarded as audience-participation, experience-simulating movies. An interactive movie consists of the following elements:

- (1) An interactive story that develops differently depending on the interaction of the audience;
- (2) An audience that becomes the main character and experiences the world created by the interactive story,
- (3) Characters who interact with the main character(audience) in the story.

Interactive movies have the following functions.

- (1) The use of CG technology and the generation of three-dimensional imagery create a virtual reality that the audience perceives as the actual surroundings.
- (2) The audience can enjoy the story development by interacting with the characters in cyberspace through talking and gesturing.

#### 3.2 Comparison with Other Media

The following is a comparison of interactive movies with other media.

##### (1) Telecommunications

Research has been conducted to reproduce a three-dimensional visual image of a person and a background at a distance displayed, so that another person can communicate with that person in a seemingly face-to-face situation. An example of the results achieved by this research is a teleconferencing system with true-to-life presence[5]. This communication concept is implemented as an advanced version of a teleconference and thus does not include a story development feature.

##### (2) Movies

Movies use dynamic images and sound to provide strong input to the human visual and auditory sensory systems and draw audiences into a cyberspace. The rapid advance of CG technology in recent years has enabled the production of very realistic images of extraordinary worlds and phenomena that do

not exist or happen in the real world. There have been attempts to add an interactive function to those movies. However, such efforts were limited to a primitive level: several different story developments were prepared, and an audience selected from available options.

##### (3) Video games

Video games, particularly role playing games (RPGs), provide a reality similar to that created by novels but in a game format. In an RPG, a basic story is preset. The player controls the story development by manipulating the main character. In a sense, video games have many similarities to interactive movies. In a video game, however, interactions are carried out by operating buttons, and this is a major difference from the natural interaction that an interactive movie allows an audience to engage in.

##### (4) Other media

Various experiments have been conducted to create a cyberspace in which an audience can interact with a virtual reality or with movie characters. These include computer-generated characters[2][9] that interact with people through gestures and emotional expressions, and interactive art[8]. However, the interaction provided by these media is short in time duration, and no story is constructed.

#### 3.3 Configuration of the First Prototype System

Based on the concept described above, we developed our first prototype system[4]. The following is a brief outline of this system.

##### 3.3.1 Software

Figure 1 shows the software configuration of the system. The interactive story consists of a collection of various scenes and

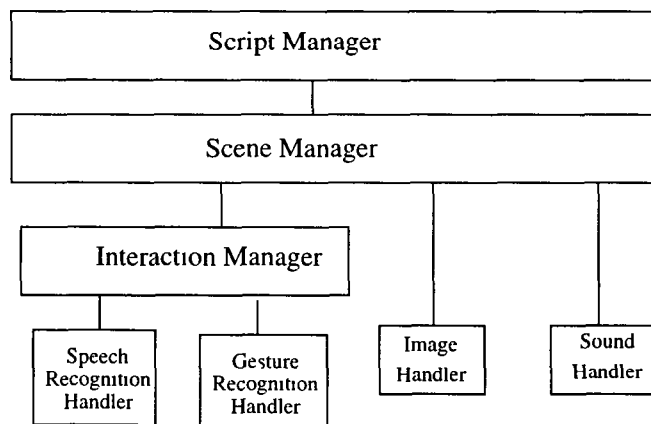


Fig. 1 Software configuration of the first prototype system

a state transition network between the scenes.

The script manager stores the data of the state transition network and controls the scene transition according to the interaction result. The scene manager contains descriptive data of individual scenes. The descriptive data includes background scene/music with their starting times, character animations/dialogs with their starting times, and so on. The scene manager generates each scene by referring to the descriptive data of the scene specified by the script manager. The interaction manager is under the control of the script manager and scene manager, and it manages the interaction in each scene. An interaction is achieved by the speech recognition function and gesture recognition function. The handlers are controlled by the scene manager and interaction manager. They control various input and output functions. The speech recognition handler controls the speech recognition function. The speech recognition program is equipped with speaker-independent and continuous speech recognition functions based on HMM[7]. The gesture recognition handler controls the gesture recognition function. Gesture recognition software detects several characteristic points of a human figure based on an image captured by a camera[10]. The image handler controls the output of visual images, such as background images and character animation. Finally, the sound handler controls the output of sounds, such as background music, sound effects and character dialogs.

### 3.3.2 Hardware

Figure 2 shows the hardware configuration. The system consists of an image output subsystem, a speech recognition subsystem, a gesture recognition subsystem and a sound output subsystem.

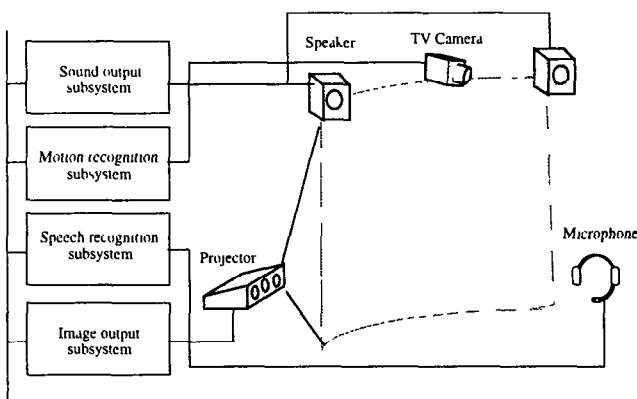


Fig. 2 Hardware configuration of the first prototype system

In the image output subsystem, a high-speed CG-generating workstation is used for the visual image output. Character images and background CG images are stored as CG animation data and used for the real-time generation of CG images. The speech recognition operations are executed by a single workstation. The gesture recognition function is also executed by a workstation. The sound output subsystem consists of several PCs. Background music, sound effects and character dialogs are produced simultaneously by this subsystem.

### 3.4 Evaluation and Problems

We tested the first prototype system with approximately 50 people during a half-year period from the time the system was developed. These people included researchers at ATR and visitors to our research facility. Although we plan to later conduct a more structured evaluation, we were successful in obtaining the participants' comments. Based on their comments, we evaluated the system and identified areas for further research, as summarized below

#### (1) Participation in cyberspace

##### a) Number of participants

In the first system, the basic concept was a story with just one player acting the role of the hero. However, the first system lacks certain functions needed for the story to take place in cyberspace since cyberspace will be created over a network and will require the story to develop from not just one player, but from among several players participating at the same time.

##### b) Presence or absence of an avatar

When players participate in a story in cyberspace, the question then becomes whether to give the players a persona (avatar). The first system did not use avatars because the one player always acted out the leading role. Even so, players never really felt that they were active participants with the system because aspects such as clothing never matched the situation in cyberspace (being in the past or on an unknown planet). To that end, we conducted research in areas such as a virtual Kabuki system[6] that uses avatars to change players into forms (a Kabuki actor in this case) that fit in cyberspace, and the results suggested the benefits of using avatars.

#### (2) Interaction

##### a) Frequency of interaction

Interaction in the first system was generally limited to change points in the story, so the story progressed linearly along a predetermined course like a movie except at these change points. Even though there are certain advantages to this tech-

nique, such as being able to use movie story development technology and expertise, the disadvantage of fixed story elements created as in a movie is that the player seems to end up a spectator who finds it difficult to participate interactively at points where interaction is clearly required. The limited opportunities for interaction create other drawbacks for the player, such as having little to distinguish the experience from watching a movie and having a very limited sense of involvement.

#### b) Types of interaction

The interaction technologies used in the first system were voice and gesture recognition. However, only exaggerated gestures rather than minute gestures could be recognized because of low lighting in the area where the system played. As a result, the system ran almost exclusively on voice recognition and players were limited to available modalities that allowed only simple interaction.

### 4. DESCRIPTION OF THE SECOND SYSTEM

#### 4.1 Improvement

The following points were used to improve the second system as described below.

##### (1) Participation in cyberspace

###### a) System for multiple players

Our initial effort to develop a system for multiple players allowed two players to participate in cyberspace in the development of a story. The ultimate goal was to create a multi-player system operating across a network, but the first step in the present study was to develop a prototype multi-player system with two systems connected by a LAN.

###### b) Avatar representation

We used a system that showed avatars as alter egos of the players on screen. There were several advantages to this system, as outlined below.

**System for representing the avatar:** The relationship of the player to the avatar and the relationship of the avatar to other characters in the movie can be controlled in various ways by changing the representational form of the avatar.

**System for controlling the avatar:** The basic control system inputs player movement by using magnetic sensors and uses that movement to map avatar movement in a motion capture system. Giving autonomy to avatar movement enables complex movement that combines autonomous avatar movement with player movement. By varying the proportion of each movement with time and circumstance, player movement can

be used directly or the player can introduce desired movement that together add diversity and depth to the relationship of the player to cyberspace.

#### (2) Interaction

##### a) Introduction of interaction any time

To increase the frequency of interaction between the participants and the system, we devised a way for players to interact with cyberspace residents at any point in time. Basically, these impromptu interactions, called story unconscious interaction (SUI), occur between the players and characters and generally do not affect story development. On the other hand, there are sometimes interactions that affects story development. This kind of interaction, called story conscious interaction (SCI), occurs at branch points in the story, and the results from such an interaction determine the future development of the story.

##### b) Introduction of multimodal interaction

The following interactive functions were added to the primary interactive function of voice recognition.

**Emotion recognition:** To realize interaction at any time, an emotion recognition capability was introduced. When players utter spontaneous utterances, the characters react by using their own utterances and animations according to the emotion recognition result. Emotion recognition is achieved by using a neural-network-based algorithm[9].

**Motion capture:** We used a motion capture system based on magnetic sensors attached to applicable parts of the player's body in order to reflect player movement in avatar movement. Data from the magnetic sensors is input into the system in order to move the computer graphic avatar so that players get the feeling they are controlling the movement of their avatar. This provides another form of interaction at any time.

**Gesture recognition:** We captured motion with magnetic sen-

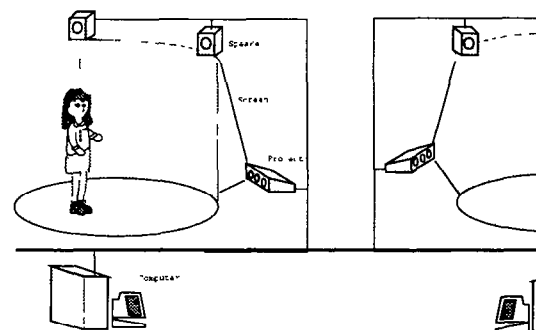


Fig. 3 Appearance of the second system

sors and used an HMM to process data from the sensors in order to recognize 3-D gestures, minute gestures and gestures under low-light conditions. Gesture recognition results are used for SCI.

Fig. 3 shows an overview of the second system equipped with these functions.

## 4.2 Software System Structure

Figure 4 shows the structure of the software used in the second system.

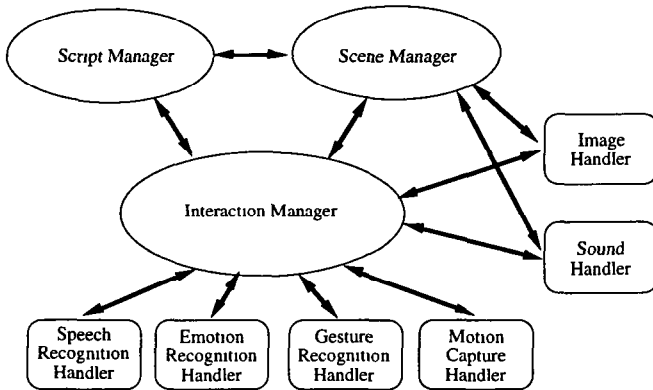


Fig. 4 Software configuration of the second system

### (1) System structure concept

While the first system stressed story development, the second system had to achieve a good balance between story development and impromptu interaction by incorporating the concept of interaction at any time. This required building a distributed control system instead of a top-down system structure.

There is some variety in the architectures available for distributed control systems, but we chose to use an action selection network [3] that sends and receives activation levels among multiple nodes. These levels activate nodes and trigger processes associated with the nodes at a point beyond the activation level threshold.

### (2) Script manager

The role of the script manager is to control transitions between scenes, just as it did with the first system. An interactive story consists of various kinds of scenes and transitions among scenes. The functions of the script manager are to define the elements of each scene and to control scene transitions based on an infinite automaton (Fig. 5). The transition from a single scene to one of possible consecutive scenes is decided based on the SCI result sent from the scene manager.

### (3) Scene manager

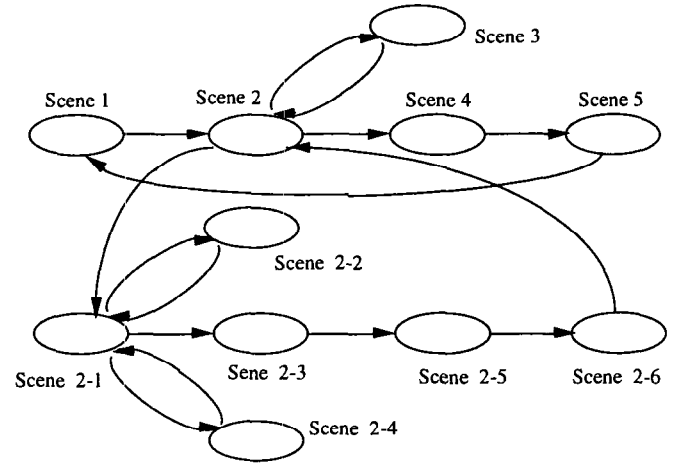


Fig. 5 Scene transition network

The scene manager controls the scene script as well as the progress of the story in a scene. Action related to the progress of the story in a scene is called an event, and event transitions are controlled by the scene manager. Events for each scene derive from the following elements.

- 1) Scene images
- 2) Background music
- 3) Sound effects
- 4) Character animation and character speech
- 5) Player and character interaction

The script for each scene is stored ahead of time in an event network, and the scene manager generates each scene based on data from the script manager via a script in the format shown in Fig. 6.

The timing for transition from one event to the next was controlled by the scene manager in the first system, but absolute time cannot be controlled in the second system because it incorporates the concept of interaction at any time. However, relative time or time order can be controlled in the second sys-

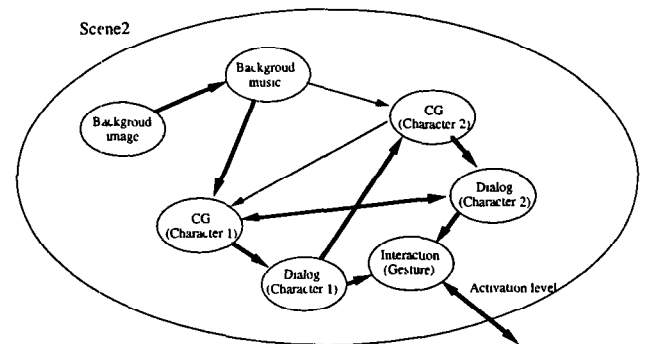


Fig. 6 An example of scene data

tem, so the action selection network was applied here as well. The following describes how this works.

- 1) Activation levels are sent or exchanged among events as well as external events.
- 2) An event activates when the cumulative activation level exceeds the threshold.
- 3) On activation of an event, a predetermined action corresponding to the event occurs. At the same time, activation levels are sent to other events, and the activation level for the activating event is reset. The order of events can be preset, and variation as well as ambiguity can be introduced into the order of events by predetermining the direction that activation levels are sent and the strength of activation levels.
- 4) Interaction manager

The interaction manager is the most critical aspect for achieving anytime interaction. Figure 6 shows the structure of the interaction manager. The basis for anytime interaction is a

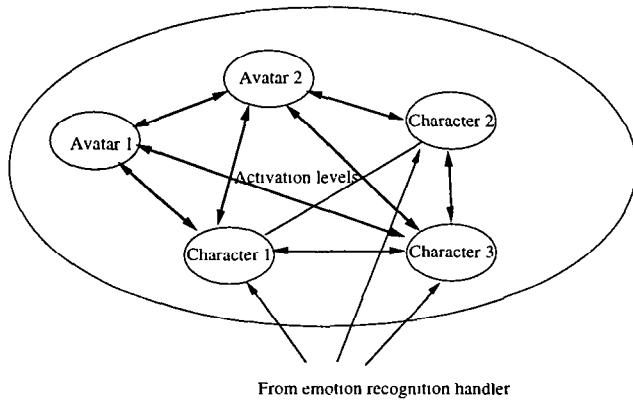


Fig. 7 Structure of interaction manager

structure where each character (including the player's avatar) is allotted an emotional state, and interaction input from the player as well as interaction between the characters determines the emotional state as well as the response to that emotional state for each character. Some leeway is given to how a response is expressed depending on the character's personality and circumstances. The interaction manager is designed based on the concepts outlined below.

- 1) Defining an emotional state

The state and intensity of a player's ( $i = 1, 2, \dots$ ) emotion at time  $T$  is defined as follows:

$$Ep(i, T), sp(i, T) \text{ where } sp(i, T) = 0 \text{ or } 1$$

(0 indicates no input and 1 indicates an input.)

Similarly, the state and intensity of an object's ( $i = 1, 2, \dots$ ) emo-

tion at time  $T$  is defined as follows:

$$Eo(i, T), so(i, T)$$

- 2) Defining the emotional state of an object

For the sake of simplicity, the emotional state of an object is determined by the emotional state when player interaction results from emotion recognition.

$$\{Ep(i, T)\} \rightarrow \{Eo(j, T + 1)\}$$

Activation levels are sent to each object when emotion recognition results are input.

$$sp(i, T) \rightarrow sp(i, j, T)$$

Where  $sp(i, j, T)$  is the activation level sent to object  $j$  when the emotion of player  $i$  is recognized. The activation level for object  $j$  is the total of all activation levels received by the object.

$$so(j, T + 1) = \sum sp(i, j, T)$$

- 3) Exhibiting actionAn object that exceeds the activation threshold performs action  $Ao(i, T)$  based on an emotional state. More specifically, action is a character's movement and speech that is a reaction to the emotional state of the player. At the same time, activation levels  $so(i, j, T)$  are sent to other objects.

$$\text{if } so(i, T) > TH_i$$

$$\text{then } Eo(i, T) \rightarrow Ao(i, T), Eo(i, T) \rightarrow so(i, j, T)$$

$$so(j, T + 1) = \sum so(i, j, T)$$

This mechanism creates interaction between objects, and enables more diverse interaction than simple interaction with a one-to-one correspondence between emotion recognition results and object reactions.

## 4.3 Hardware System Structure

Figure 8 shows the second system's hardware structure, comprised of image output, voice and emotion recognition, gesture recognition and sound output subsystems.

- (1) Image output subsystem

Two workstations (Onyx Infinite Reality and Indigo 2 Impact) capable of generating computer graphics at high speed are used to output images. The Onyx workstation is used to run the script manager, scene manager, interaction manager and all image output software. Character images are stored on the workstations ahead of time in the form of computer graphic animation data in order to generate computer graphics in real time. Background computer graphic images are also stored as digital data so background images can be generated in real time. Some background images are real photographic images stored on an external laser disc. The multiple character computer graphics, background computer graphics and background

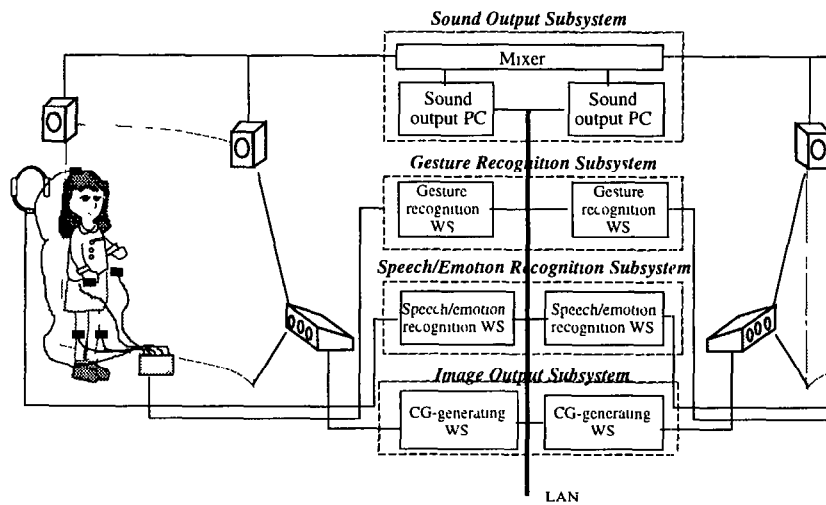


Fig. 8 Hardware configuration of the second system

photographic images are processed simultaneously through video boards on both the Onyx and Indigo 2 workstations.

Computer graphics are displayed in 3-D for more realistic images, and a curved screen is used to envelop the player with images and immerse the player in the interactive movie world. Image data for the left and right eye, created on the workstations ahead of time, are integrated by stereoscopic vision control and projected on a curved screen with two projectors (Fig. 9.). On the Indigo 2 end, however, images are output on an ordinary large-screen display without stereoscopic vision because of processing speed

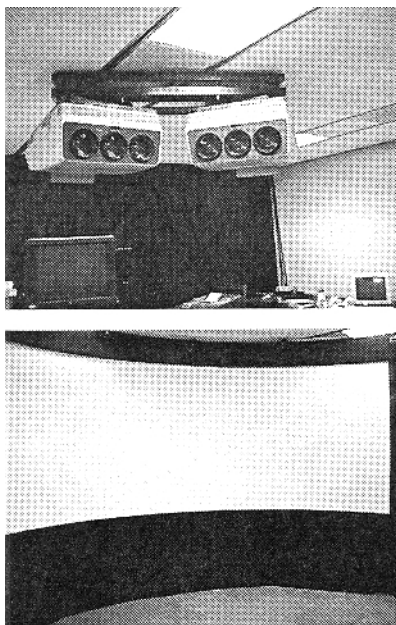


Fig. 9 Projectors and arch screen

#### (2) Voice and emotion recognition subsystem

Voice and emotion are recognized with two workstations (Sun SS20s) that also run the voice and emotion recognition handlers. Voice input via microphone is converted from analog to digital by the sound board built into the Sun workstation, and recognition software on the workstation is used to recognize voices and emotions. For the recognition of meaning, speaker-independent speech recognition algorithm based on HMM is adopted also in the second system. Each workstation processes voice input from one player.

#### (3) Gesture recognition subsystem

Gestures are recognized with two SGI Indy workstations that run the gesture recognition handlers. Each workstation takes output from magnetic sensors attached to a player and uses that data output for both controlling the avatar and recognizing gestures.

#### (4) Sound output subsystem

The sound output subsystem is comprised of several personal computers because background music, sound effects and speech for each character must be output simultaneously. Sound effects and character speech are stored as digital data that is converted from digital to analog as needed, and multiple personal computers are used to enable simultaneous digital to analog conversion of multiple channels in order to output these sounds simultaneously. Background music is stored ahead of time on an external compact disc whose output is also controlled by the personal computer. The multiple channel sound outputs are mixed and output with a mixer (Yamaha 02R) that can be controlled by computer.

## 5. EXAMPLE OF INTERACTIVE STORY PRODUCTION

### 5.1 An interactive story

We have produced an interactive story based on the previously described system. We selected "Romeo and Juliet" by Shakespeare as the base story for the following reasons.

a) There are two main characters, Romeo and Juliet, in the story and, therefore, this story supplies a good example of multi-person participation.

b) "Romeo and Juliet" is a very well known story, and people have a strong desire to act out the role of hero or heroine. Therefore, it is expected that people can easily get involved in the movie world and experience the story

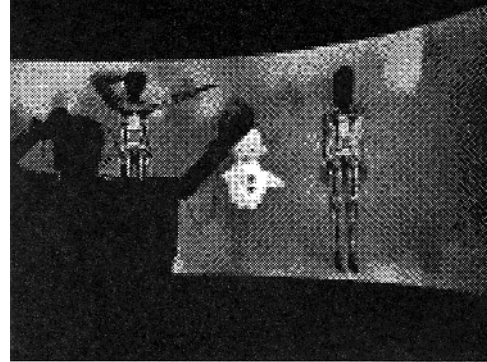
The main plot of the story is as follows. After their tragic suicide their souls are sent to Hades, where they find that they have totally lost their memory. Then they start their journey to rediscover who they are and what their relationship is. With various kinds of experiences and with the help and guidance of characters in Hades, they gradually find themselves again and finally go back to the real world

### 5.2 Interaction

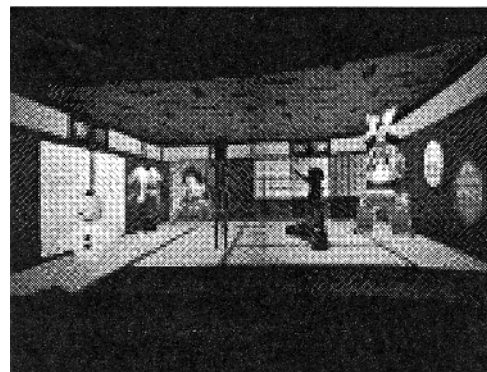
There are two participants. one plays the role of Romeo and the other Juliet. The two subsystems are located in two separate rooms and connected by a LAN. Each participant stands in front of the screen of his/her respective system wearing specially designed clothes to which magnetic sensors and microphones are attached. In the case of Romeo, the participant wears a 3-D LCD-shutter glass and can enjoy 3-D scenes. Their avatars are on the screen and move according to their actions. They can also communicate by voice. Basically, the system controls the progress of the story with character animations and character dialogues. Depending on the voice and gesture reactions of the participants, the story moves on. Furthermore, as is described before, interaction at any time is possible. When the participants utter, the characters react, according to the emotion recognition results. Consequently, depending on the frequency of the participants' interaction, this system can go anywhere between story-dominant operation and impromptu interaction-dominant operation. Figure 10 illustrates typical interactions between the audience and the system.

## 6. CONCLUSIONS

Interactive movies are a new type of media that integrates vari-



(a) "Romeo" controls his avatar



(a) "Romeo" tries to touch objects in Japanese antique shop

Fig. 10 Examples of interactions between participants and the system

ous media types including telecommunications, movies, and video games. In interactive movies, people enter cyberspace and enjoy the development of a story there by interacting with the characters in the story. In this paper, we first explained the concept of interactive movies and briefly explained our first prototype system. Based on an evaluation of this system, we identified several problems in the system that needed to be improved. One is the lack of frequent interactions and the other is single-person participation. To improve these deficiencies, we are developing a second system. We explained two significant improvements incorporated into the second system: interaction at any time and two-person participation through a network. We described the software and hardware configuration of the second system while putting emphasis on these improvements. We also described two interaction tech-



nologies that are introduced for the first time in the second system and showed that are the keys to realizing these improvements. Finally, we illustrated the operation of our Second System with the example of our interactive production of "Romeo and Juliet".

## 7. REFERENCES

- [1] Ceram, C. W. *Eine Archäologie des Kinos*. Rowohlt Verlag, Hamburg, 1965.
- [2] Maes, P., et al. The ALIVE system: Full-body Interaction with Autonomous Agents. in *Proceedings of the Computer Animation'95 Conference*, 1995.
- [3] Maes, P. How to do the right thing. *Connection Science*, Vol.1, No.3, 291-323, 1989.
- [4] Nakatsu, R., and Tosa, N. Toward the Realization of Interactive movies - Inter Communication Theater: Concept and System. in *Proceedings of the International Conference on Multimedia Computing and Systems'97*, 71-77, 1997.
- [5] Noma, H., et al. Multi-Point Virtual Space Teleconference System. *IEICE Trans. Commun* , Vol. E78-B, No. 7, 1996.
- [6] Ohya, J., et al. Virtual Kabuki Theater: Towards the Realization of Human Metamorphosis Systems. in *Proceedings of 5th IEEE Workshop on Robot and Human Communication (RO-MAN'96)*, 416-421, 1996.
- [7] Shimizu, T., et al. Spontaneous Dialogue Speech Recognition Using Cross-Word Context Constrained Word Graph. in *Proceedings of ICASSP'96*, Vol. 1, 145-148, 1996.
- [8] Sommerer, C., and Mignonneau, L. A-Volve: An Evolutionary Artificial Life Environment. in *Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems*, 167-175. 1997.
- [9] Tosa, N., and Nakatsu, R. Life-like Communication Agent - Emotion Sensing Character 'MIC' and Feeling Session Character 'MUSE' -. in *Proceedings of the International Conference on Multi-media Computing and Systems*, 12-19, 1996.
- [10] Wren, C. R., et al. Pfinder: Real-Time Tracking of the Human Body. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, 780-785, 1997.