Creation of Virtual Theater
- Interactive Poem and Play Cinema -

Naoko Tosa and Ryohei Nakatsu
ATR Media Integration and Communications Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan
{tosa, nakatsu}@mic.atr.co.jp, http://www.mic.atr.co.jp/

Abstract

In the conventional theater, the audience enjoy the ongoing of play or movie as the observers. By implementing interaction capabilities to the conventional theatrical media, however, there is a possibility that a new type of media emerges. We call this new type of media as "Virtual Theater," and as examples of virtual theater we propose "Interactive Poem" and "Play Cinema." In the interactive poem system, a human and a computer agent create a poetic world by exchanging poetic phrases, thus realizing sensitivity-based communications between computers and humans. In Play Cinema, people enter cyberspace and enjoy the development of a story by interacting with computer characters in the story. In this paper we first explain the concept of Interactive Poem and describes the details of the system we have developed. Then we explain the concept of of Play Cinema and describe the construction of the system we are currently developing.

1. Interactive Poem

1.1 Outline of Interactive Poem

Concept

Humans seek meaning through conversation. At times, as extreme examples, we even find ourselves talking to dolls and objects. When we do this, there is no warmth nor richness when our reactions switch like the character "Tamagochi".

For our part at ATR, we have tried to consider interaction with computer characters that can handle the meanings of words[1][2]. Also, many research have been carried out to realize human-like computer characters[3][4][5]. However, it has been impossible to carry out inspirational Japanese conversations on computers in real time. Then, one day, when we were thinking about setting up a framework of conditions able to serve as a foundation, we hit upon the idea of linked verses from our ancient culture. "Renga" is generated by multiple people as a combination of short Japanese poems such as "Waka" or "Haiku" which were created in ancient era and have been used as medium to express Japanese spiritual emotions. Therefore, we worked our way toward thinking about having computer poets and humans construct improvised poems in the form of linked verses. Fundamentally, poems are what individual poets utilize to construct their worlds using the "power of words" to express their messages. By reading and listening to these words, all of us are able to enjoy the worlds of the poets. Interactive poems allow us not only to enjoy the worlds of poets. Individual personalities are expected to emerge by having humans work with computers to actively make poems, through unex-

pected happenings or by chance, or depending on what persons are working with the computers to make poems. Therefore, these poems create openings through the interaction process. In other words, a medium for communications through "Japanese" is in place with the adoption of a function for conversation.

Experience with Interactive Poem

The face of the music goddess "MUSE" in Greek mythology appears on a large screen. MUSE, as if singing in complete unison, creates poems while conversing with a person. This computer character speaks short poetic words with emotions to the person. By hearing these words, the person is able to enter the world of that poem, and at the same time, he or she is also able to speak to MUSE with poetic words. Through this process of exchanging poetic words, the interactive poem allows the user and computer to work together to build the world of an improvised poem filled with inspiration, feeling, and "Japanese spiritual Emotion" (Figure 1).

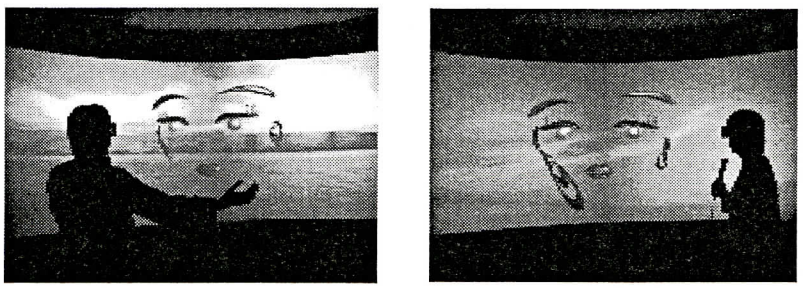


Fig. 1. Example of interaction between Muse and the audience

1.2 Software Configuration

The system used to create the interactive poem consists of four main units: system control, speech recognition, computer graphics generation and speech output (Fig. 2).

The system control unit manages behavior of the whole system by utilizing the inter-

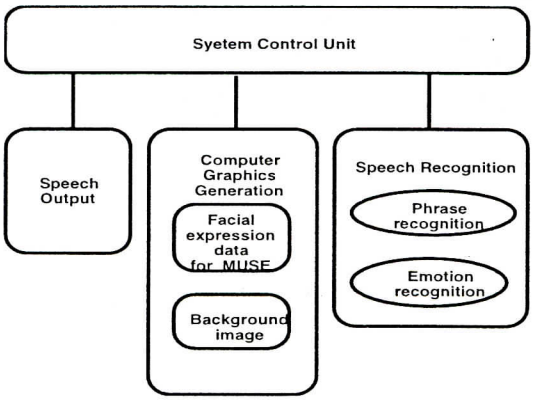


Fig. 2. Software configuration of the Interactive Poem system

active poem database. In this system, the most important issue is constructing the interactive poem, so we must first explain how the interactive poem database is constructed. A conventional poem is considered a sequence of poetic phrases. In other words, the basic construction of a conventional poem can be expressed by a simple state-transition network where each phrase corresponds to a given state, and for each state there is only one successive state (Fig. 3).

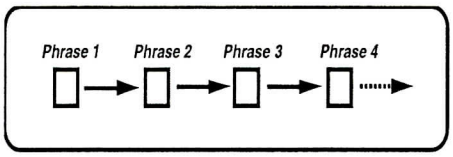


Fig. 3. Construction of conventional poem

The basic form of the interactive poem is expressed by this simple transition network, but it differs from a conventional poem in that phrases uttered by the computer agent and phrases uttered by a participant appear in turn. This corresponds to a simple interaction where the computer agent and a participant alternately read a predetermined sequence of poetic phrases (Fig. 4-a).

To introduce improvisational interaction into our system, we modified this simple transition so that multiple phrases are connected to each phrase of the computer agent (Fig. 4-b). These phrases are carefully created and chosen by taking into account how

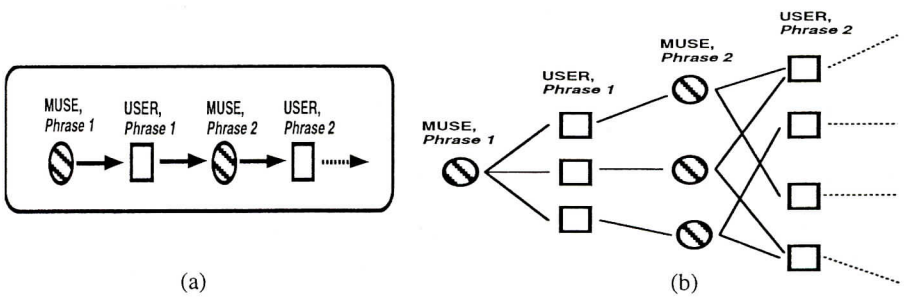


Fig. 4. Construction of interactive poem

well their rhythms are formed and the meaning of each phrase. This transition network is stored in the interactive poem database and used to control the whole process.

The speech recognition unit has two different speech recognition functions: phrase recognition and emotion recognition. To recognize each phrase uttered by a participant, we have adopted HMM (hidden Markov model) based speaker-independent speech recognition technology[9]. Each phrase to be uttered is represented in the form of a phoneme sequence and is stored in the lexicon. To simultaneously detect the emotional state of a participant, an emotion recognition function is introduced. A neural network architecture has been adopted as the basic architecture for emotion recognition[2]. This neural network is trained with the utterances of many speakers to express the eight emotional states of joy, happiness, anger, fear, teasing, disgust, disappointment and emotionless. As such, speaker-independent and content-independent emotion recognition

is realized.

Reaction of the computer agent to utterances of a participant is expressed through her speech and by images. In the speech output unit, speech data for each phrase to be uttered by the computer agent is digitally stored and generated when necessary.

The computer graphics generation unit controls image reaction of the computer agent. Image reaction consists of two kinds of images: facial expressions for the computer agent "MUSE" and various scenes. The facial expressions of MUSE express her reactions to the emotional state of a participant. These images are represented by key frame animations, each of which corresponds to the eight emotions (Fig. 5). To express the atmosphere of the interactive poem, several kinds of scenes are digitally stored. Each scene image corresponds to a group of states in the transition network, and each correspondence is carefully determined in advance.

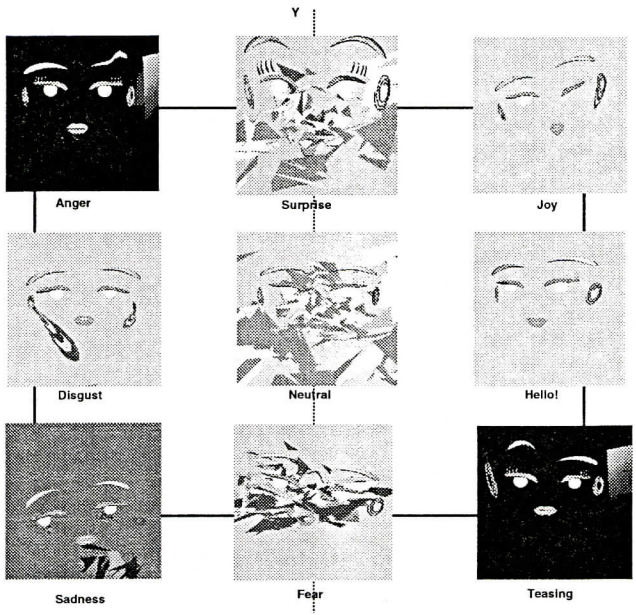


Fig. 5. Muse's emotional expression

1.3 Hardware Configuration

Figure 6 illustrates the hardware configuration of the system. This mainly consists of several workstations and a PC: a workstation for computer graphics generation, a workstation for both system control and phrase recognition, a workstation for emotion recognition, and a PC for speech output. For the participant's convenience, optional phrases that may be uttered following an utterance of MUSE appear on the display. The participant can choose one of these phrases based on their feelings and sensitivity. Or they can create their own poetic phrase. In this case, the phrase recognition function selects the preexisting phrase that most resembles the uttered phrase. Therefore, the participant will feel as if the interactive poem process continues in a natural way.

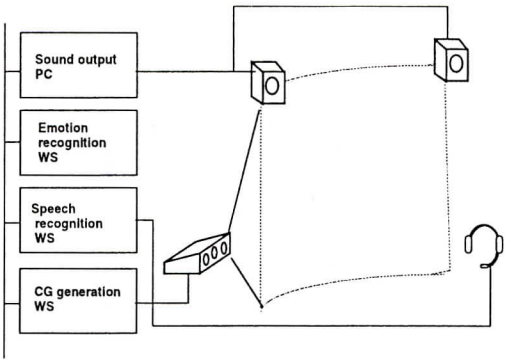


Fig. 6. Hardware configuration of the Interactive Poem system

1.4 Interactions

The interaction mechanism operates as follows.

(1) When MUSE utters a phrase, the recognition process is activated. A participant then utters a phrase and it is recognized by the phrase recognition function, which uses the lexicon subset corresponding to the next set of phrases in the transition network. At the same time, emotion contained in the utterance is recognized by the emotion recognition function.

(2) Based on information pertaining to recognition and the transition network, the system's reaction is decided. The facial expression of MUSE changes according to the results of emotion recognition, and the phrase MUSE utters is based on the results of phrase recognition and the transition network. The background scene changes as the transitions continue.

(3) In the above stated manner, poetic phrases between MUSE and the participant are consecutively produced.

2. Play Cinema

2.1 Outline of Play Cinema

Ever since the Lumiere brothers created cinematography at the end of the 19th century[6], motion pictures have undergone various advances in both technology and content. Today, motion pictures, or movies, have established themselves as a composite art form that serves a wide range of cultural needs extending from fine art to mass entertainment. However, conventional movies unilaterally present predetermined scenes and story settings, so audiences take no part in them and make no choices in story development. On the other hand, the use of interaction technology makes it possible for the viewer to "become" the main character in a movie and enjoy a first-hand experience. We believe that this approach would allow producers to explore the possibilities of a new class of movies.

From this viewpoint, we have been conducting research on "Play Cinema" production by applying interaction technology to conventional movie making techniques. As an

initial step in creating a new type of movie, we have produced a prototype system[7]. Based on this system, we are currently developing a second prototype system with many improvements.

Concept

Compared with existing media, Play Cinema can be regarded as audience-participation, experience-simulating movies. A Play Cinema system 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.

Configuration of the first prototype system

Based on the concept described above, we developed our first prototype system[7]. Figure 7shows the software configuration of the system. The interactive story consists of a collection of various scenes and a state-transition network between the scenes.

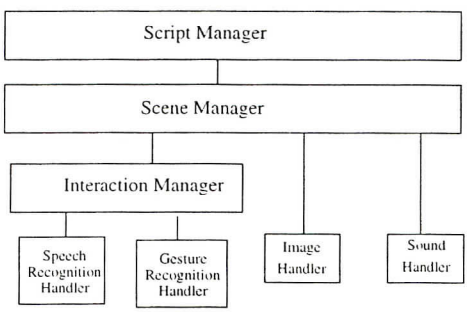


Fig. 7. Software configuration of the first prototype system

The script manager stores the data of the state-transition networks and controls the scene transition according to the interaction result. The scene manager contains descriptive data of individual scenes and 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. The handlers are controlled by the scene manager and interaction manager. They control various input and output functions such as speech recognition, gesture recognition, output of visual images and output of sounds.

Evaluation and problems

We tested the first prototype system with approximately 50 people during the half-year period following completion of system development. Based on their comments, we evaluated the system and identified areas for further research, as summarized below.

(1) Number of participants

The basic concept of the first system was a story with just one player acting the role of

hero. However, the first system lacked the multi-user functions needed for the story to take place in cyberspace, though cyberspace will be created over a network and will require the story to develop from not just one player but from the interactions of several players participating at the same time.

(2) 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. There are certain advantages to this technique, such as being able to use the story development techniques and expertise accumulated by skilled cinematographers. However, the disadvantage of using fixed story elements created in the same way as for a conventional 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.

2.2 Improvement

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

(1) 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 the development of a prototype multi-player system consisting of two systems connected by a LAN.

(2) Introduction of interaction at 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 do affect story development. This kind of interaction, called story conscious interaction (SCI), occurs at branch points in the story, and the results of such an interactions determine the future development of the story.

(3) Other improvements

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.

Motion capture: We introduced a motion capture system based on magnetic sensors. There are two major reasons for such a system. One is to show avatars as alter egos of the players on screen, thus giving the players the feeling that they are really active participants with the system. The other is to improve gesture recognition. The first system's gesture recognition based on images obtained by a camera was ineffective due to low light. Therefore, we wanted to use motion capture data for gesture recognition.

2.3 Software Configuration

Figure 8 shows the structure of the software used in the second system.

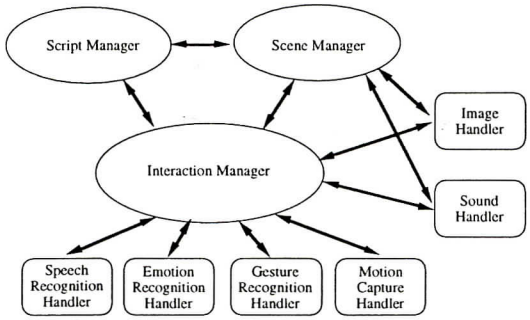


Fig. 8. Software configuration of the second system

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 a variety of architectures available for distributed control systems, but we chose to use an action selection network [8] 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.

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. 9). The transition from a single scene to one of several possible subsequent scenes is decided based on the SCI result sent from the scene manager.

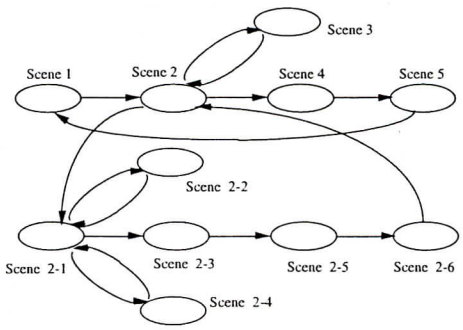


Fig. 9. Sene transition network

Scene manager

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 consist of scene images, background music, sound effects, character animation and character speech, and player and character interaction

The script for each scene is pre-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.

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 and time order can be controlled in the second system, 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.

Interaction manager

The interaction manager is the most critical component for achieving interaction at any time. Figure 10 shows the structure of the interaction manager. The basis of

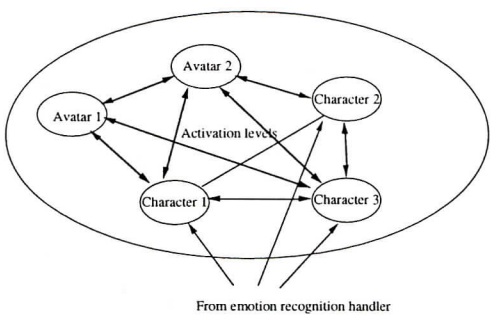


Fig. 10. Structure of interaction manager

interaction at any time is a structure that allots each character (including the player's avatar) an emotional state, and interaction input from the player along with interaction between the characters determine 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...) emotion at time T is defined as

Ep (i, T), sp(i, T) where sp(i, T) = 0 or 1
(0 indicates no input and 1 indicates an input).

Similarly, the state and intensity of an object's (i = 1, 2...) emotion at time T is defined as

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

(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:

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

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

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

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:

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

(3) Exhibiting action

An 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 as a reaction to the emotional state of the player. At the same time, activation levels so(i, j, T) are sent to other objects:

$$\text{if } so(i, T) > THi$$
$$\text{then } Eo(i, T) \longrightarrow Ao(i, T),\ Eo(i, T) \longrightarrow so(i, j, T)$$
$$so(j, T + 1) = \Sigma\ so(i, j, T).$$

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.

2.4 Hardware Configuration

Figure 11 shows the second system's hardware structure, composed of image output, voice and emotion recognition, gesture recognition and sound output subsystems.

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 pre-stored on the workstations 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

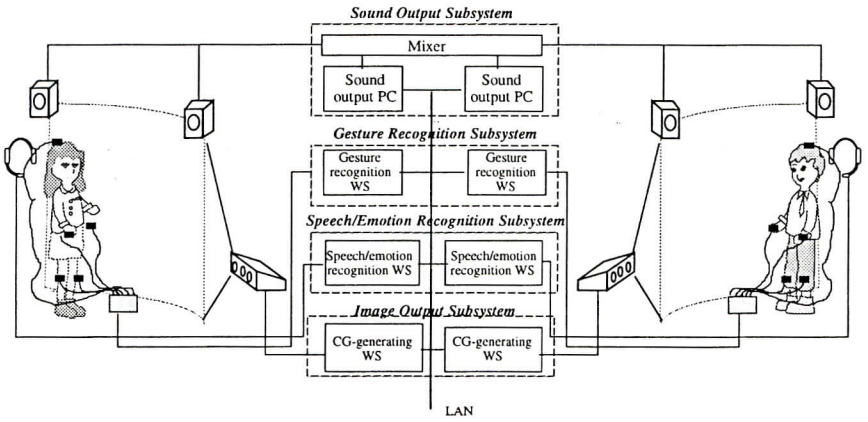


Fig. 11. Hardware configuration of the second system

generated in real time. Some background images are photographic images of real scenery stored on an external laser disc. The multiple character computer graphics, background computer graphics and background 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 him or her in the interactive movie world. Image data for the left and right eye, previously created on the workstations, are integrated by stereoscopic vision control and projected onto a curved screen by two projectors. On the Indigo 2 end, however, images are output on an ordinary large-screen display without stereoscopic vision because of processing speed.

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, a speaker-independent speech recognition algorithm based on HMM is adopted[4]. Emotion recognition is achieved by using a neural-network-based algorithm[5]. Each workstation processes voice input from one player.

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.

Sound output subsystem

The sound output subsystem uses 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 are 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 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 .

2.5 Example of Interactive Story Production

An interactive story

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

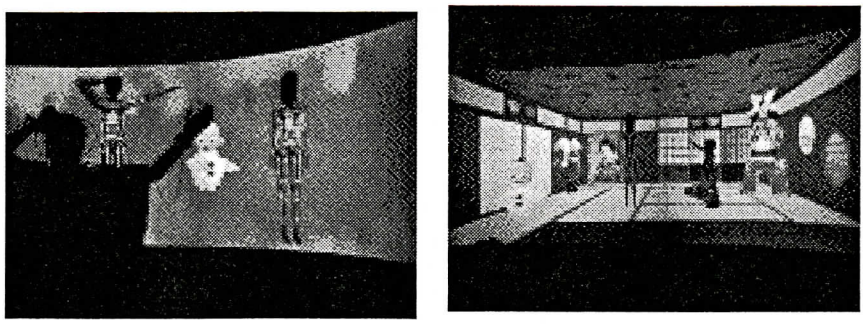
(1) There are two main characters in the story and, therefore, this supplies a good example of multi-person participation.

(2) "Romeo and Juliet" is a very well known story, and people have a strong desire to act out the role of hero or heroin. 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 the lovers' 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.

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 is possible at any time. When the participants utter, the characters react according to the



(a) "Romeo" controls his avatar

(b) "Romeo" tries to touch object in Japanese curiosity shop

Fig. 12. Examples of interaction between participants and system

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 12 illustrates typical interactions between the participants and the system.

3. Conclusion

In this paper we have proposed a concept of "Virtual Theater." Virtual Theater is a new type of media that implements interaction capabilities to the conventional theatrical media such as play or cinema. In these media, audience are only the observers of the play or cinema. This means that their role is rather passive. In the case of Virtual Theater, as the audience can interact with the play/cinema world directly and can change the story development in these worlds. Therefore they are not the passive observers but active participants to the events of the virtual world. As examples of Virtual Theater, we have proposed two systems; "Interactive Poem" and "Play Cinema."

Interactive poem is a new type of poem created by a participant and a computer agent collaborating in a poetic world full of inspiration, emotion and sensitivity. In Play Cinema people enter cyberspace and enjoy the development of a story by interacting with computer characters in the story. We have already developed our first prototype system. Based on an evaluation of this system, we are developing a second system where several improvements are incorporated.

Both of these system were realized by the collaboration between an artist and an engineer. By integrating these two different types of skills and talents, we could produce new system that can be considered both new interactive art media and new sensitivity-based interaction system.

References

[1] N. Tosa et al., "Neuro-Character," AAAI'94 Workshop, AI and A-Life and Entertainment (1994).

[2] N.Tosa and R. Nakatsu, "Life-like Communication Agent--Emotion Sensing Character 'MIC' and Feeling Session Character 'MUSE'," Proceedings of the International Conference on Multimedia Computing and Systems, pp.12-19 (1996).

[3] P. Maes, T. Darrell, B. Blumberg, and A. Pentland, "The ALIVE system: Full-body interaction with autonomous agents," Proc. of the Computer Animation'95 Conference (1995).

[4] K. Perlin, "Real time responsive animation with personality," IEEE Transactions on Visualization and Computer Graphics,Vol.1,No.1 pp.5-15 (1995).

[5] J. Bates, B. Loyall, and S. Reilly, "An architecture for action, emotion, and social behavior," Proceedings of the Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World (1992).

[6] C. W. Ceram, "Eine Archaologie des Kinos," Rowohlt Verlag, Hamburg (1965).

[7] R. Nakatsu and N. Tosa, "Toward the Realization of Interactive movies - Inter Communication Theater: Concept and System," Proceedings of the International Conference on Multimedia Computing and Systems'97, pp.71-77 (1997).

[8] P. Maes, "How to do the right thing," Connection Science, Vol.1, No.3, pp.291-323 (1989).

[9] T. Shimizu et al., "Spontaneous Dialogue Speech Recognition Using Cross-Word Context Constrained Word Graph," Proceedings of ICASSP'96, Vol. 1, pp. 145-148 (1996).