G. Hatano, N. Okada and H. Tanabe (Eds.)
Affective Minds

CHAPTER 19

Expression of emotion, unconsciousness with art and technology

Naoko Tosa *

ATR Media Integration and Communication Research Laboratories, 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

Abstract: Why do humans like things with human forms regardless of whether these forms depict old, young, male, or female people? Haven't humans been making forms like themselves from long ago and feeling attachment and introducing empathy toward these forms, from clay figurines to mechanical dolls? This paper focuses on the aesthetics of artificial life and communications, from the standpoint of Art and Technology. The author believes that interactive art is one type of medium that provides sympathy with communications. Interactive art can be thought of as a type of emotion and sympathy interface. It is familiar to us and provides agents or characters that can handle sensitive communications. In addition, they work on our mental states and emotional expressions, and on our character and intelligence, which means that a person can also self-create his or her own personality. On the other hand, emotion recognition technology recognizes only the surface emotion of people (Bates et al., 1992a; Maes et al., 1995; Scott Reilly and Bates, 1995). I am interested in how to recognize unconsciousness feeling by using computer-based interaction. Art is a familiar way to portray human unconscious emotion, so I tried to achieve this in interactive art with the technologies and techniques of art.

Keywords: art and technology research; interactive art; human–computer interface; computer character behavior; virtual reality; real-time computer graphics animation; multimedia; Kansei-communication; emotion recognition; neural-net work

Introduction

Human communications are composed of both verbal and non-verbal elements, and the communicators make selective or integrated use of them according to the situation (Reeves and Nass, 1996). Human–computer interaction, however, has been almost completely verbal in nature, even when creative activities are involved.

This paper discusses a method of designing a non-verbal interface centering on emotions, showing some actual cases of system construction. The research described in this paper is characterized by (1) systems developed for various levels of emotional interfaces, and (2) an artistic approach as well as an engineering approach in designing the systems.

If a non-verbal interface can be regarded as a technological extension of human emotions and sensibilities, then a variety of studies have already been made on the necessary hardware and software thanks to the development of virtual reality and other computer technologies. However, these studies have generally been aimed at reproducing the functions of the five senses we use in our daily lives. Unfortunately, the technologically reproduced senses are far from satisfactory and fail to stimulate our sensibilities because they are much less 'real' than the actual sensations we experience daily. Moreover, our explorations of sensibilities and perceptions seem to be less active now than before, probably because we are much more accustomed, or even addicted, to technology. There is also the interactive entertainment

* Correspondence to: Dr. N. Tosa, ATR Media Integration and Communication Research Laboratories, 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan. Tel.: +81 774-95-1427; Fax: +81 774-95-1408; e-mail: tosa@mic.atr.co.jp; http://www.mic.atr.co.jp/~tosa

AFFECTIVE MINDS

Edited by

G. Hatano,
N. Okada and
H. Tanabe

The 13th Toyota Conference

ELSEVIER

approach of games. But these games are still reflexive and sensory and can't arouse deep human feelings. They are little more than simple entertainment to be consumed. Adopting an artistic approach may help overcome these limitations and awaken deep emotions. Art is an original and creative activity in which the artist seeks higher levels of spiritual fulfillment by expressing his individual sensibilities with consciously trained and refined techniques. What this study sought to achieve was the construction of a system in which an artistically created computer character equipped with a model to assess emotions carries out verbal and non-verbal communications with people actively and interactively.

Importance of emotion interface

A non-verbal interface handles, as mentioned above, a vast range of objects, from those related with low-level sensations to those concerning high-level emotions, including figures, colors, movements, atmosphere, music, voice quality, expressions and sensibilities. The traditional methodologies to process sensory information can only deal with low-level sensations (Waters, 1987), and a non-verbal interface between a human and a computer should be able to handle higher-level information. However, high-level information like sensibilities allows more than one interpretation and depends to a large extent on individual elements, which makes such information difficult to handle effectively at this moment. Therefore, it is probably appropriate for now to make an interface out of the emotions that rank somewhere between the two extremes in complexity. While there are many studies in this field on emotion analysis (Morishima and Harashima, 1993), emotion recognition (Picard, 1997), and emotion synthesis (Abadjieva et al., 1993) dealing with facial expressions, few have been conducted on emotions expressed in human voice. No successful effort has yet been made to synthesize the findings in these studies into a total system incorporating verbal functions.

Trends in emotion interface research

Among the engineering studies under way abroad, the 'Cog' project (Brooks, 1996; Brooks et al., 2000), a project to create a humanoid robot led by Rodney Brooks at MIT Artificial Intelligence Laboratory, is working on an emotion generation model in cooperation with Daniel Dennett, Director of the Center for Cognitive Studies at Tufts University. The Oz Project led by Joseph Bates of Carnegie-Mellon University (Bates et al., 1992b) and the research on interactive actors led by Ken Perlin, director of New York University's Media Research Lab (Perlin, 1995), are trying to create an emotion generation model based on the movements of Disney animation characters using artificial intelligence. The Virtual Theater project led by Barbara Hayes-Roth at the Stanford University Computer Science Department (Hayes-Roth and van Gent, 1996; Isbister and Hayes-Roth, 1998) is working on an emotion generation model from the viewpoint of artificial intelligence research. In Japan, a couple of studies are focused on systems to enable a computer to analysis of the neural network recognition characteristics of 6 basic facial expression (Kobayashi and Hara, 1994) and speech dialogue with facial displays by multimodal human–computer conversation (Nagao and Takeuchi, 1994), but these systems are not able to describe the intermediate emotions that fill the gap between emotion patterns and complex emotions that emerge within a specific timeframe. They handle emotions in superficially symbolized forms and fail to encompass such delicate emotional communications like bargaining.

Importance of an artistic approach

The current work is technological research that explores human emotions and the expansion of human sensibilities and consciousness by making innovative use of artistic creativity. An artist seldom starts creating a work with a clear image of the final outcome in his mind. Consequently, the process of creating a work of art is not a process of achieving a goal but one of exploring and finding a goal. The quest is, unlike the objective, quantitative methodology used in engineering, guided by judgments based on the subjective feelings of an individual. The depth of feeling expressed and explorations of new values are praised. Since an artist presents works that reflect the times and his consciousness, the thinking and ideas behind his works can help us conceive a new sensitivity interface. A major shift from analog to digital is under way, so the time is ripe for trying to create a system to express sensibilities numerically. This is the background of my efforts to grasp the recognition and generation of emotions in this study.

The argument above may seem abstract, but it offers a very realistic approach. In her book entitled *Computers as Theatre*, theater artist and interface designer Brenda Laurel, for instance, divides the relations between information processing and users in today's computer technology into three elements: the audience watching a play, the stage, and the backstage. She argues that the principal role of computer interface technology is to place the backstage functions that support the performance on the stage, such as lighting and sound effects, out of the consciousness of the audience (Laurel, 1992). This suggests that computers can heighten human thinking and imagination when their information-processing power is combined with visual expressions that constitute a theater. This is an example of a new approach that synthesizes an engineering approach with an artistic approach. As research on the processing of sensibility information shows, various studies on the so-far untouched domain of human emotions and sensibilities are going on in fields ranging from recognition technology and artificial intelligence to the neurocomputer and brain science, exploring and analyzing diverse aspects of man's mental and spiritual activities. The achievements of these studies could be used in the future to develop a non-verbal interface. At this moment, an actual non-verbal interface can only be realized when the technology to handle emotions and an artistic approach are melded.

Implementation of the artistic approach

We are going to devise a system design that leads to the formation of a new emotional domain and complex emotions by using an artistic approach that can delineate human emotions with deep insight. Analyzing novels, poems, dramas, colors, forms and tones can provide the classification and definitions of emotional expressions, while analyzing movies and plays can provide insight into the acts and manifestations of emotional expressions that emerge within a specific timeframe.

From the standpoint of interactive art

Let us begin by considering human–computer interaction based on emotions. A typical example of an attempt to realize non-verbal interaction including emotions between a human and the computer is interactive art (Leopoldseder and Schopf, 1997). However, the traditional works of interactive art (Sommerer and Mignonneau, 1997) have often been characterized by the ideal and intellectual style used in modern art that gives the impression of machine-like coldness. The author came to the conclusion that new possibilities may be opened up if interactivity is understood in the context of the creation of a living thing that can make lively and real responses and initiate communications. Man was chosen as the object of the research because he is the creature most familiar to us and with which we have communications most frequently. A study was launched to see how realistically we could express the individuality of man's psychological conditions, emotional expressions, personality, intelligence, actions and communications with the help of technology. To detect emotions, an attempt was made to realize a non-verbal interface using emotional aspects included in voice, the most natural medium of communications for man. Then, a system unifying a non-verbal interface and a verbal interface was coupled with multi-modal interaction to incorporate complex emotions. Next, a comprehensive description of an interactive scenario within a timeframe was made to create a system for interactive movies built on emotions. In addition, a model for emotional resonance was created to discover hidden emotions in man's social communications as well as the symbolized superficial ones. Then, based on the results, an experimental work was designed in which the system could express the state of hidden relations.

Activation of empathy

Another important step in this study is the introduction of empathy. A virtual world, however splendid it may be, is only an empty facade unless it invites our empathy. People don't want to stay in such a superficial world for long. Empathy means sharing the feelings of another person. In other words, it means having a vicarious experience of the state of some-

Basic principle

Speaker independence is an important aspect of speech/emotion recognition. From a pragmatic viewpoint, a speaker-dependent emotion recognition system requires a tiresome learning stage each time a new speaker wants to use the system; therefore, it is not easy to use. Moreover, humans can understand the emotions included in speech as well as the meaning conveyed by speech, even for arbitrary speakers. In addition, content independence is indispensable for emotion recognition. In daily communication, various kinds of emotions are conveyed by the same words or sentences; mastering such nuances is the key to rich and sensitive communications among people. Therefore, by adopting neural network architecture and by introducing a training stage that uses a large number of training utterances, we have developed a speaker-independent and content-independent emotion recognition system.

Processing flow

Fig. 2 is a block diagram of the processing flow. The process mainly consists of three parts: speech processing, emotion recognition, and generation of reactions. In the speech processing part, feature parameters of input speech are extracted in real time in the feature extraction stage. Then, by observing the speech power, the period where speech exists is extracted. From the extracted speech, feature parameters are extracted and arranged as an output of the feature extraction stage. This output is fed into the emotion recognition part, where two-stage emotion recognition is carried out. In the first stage, a combination of plural neural networks, each of which is designed and trained to recognize a specific emotion in speech, receives feature parameters and carries out a recognition process. In the second stage the multiple output of the first stage is processed through a specialized logic, and the emotion recognition results are expressed as points in a two-dimensional space, in which eight emotions including neutral state are displayed according to our criteria list in Fig. 5. The result's position on the emotion plane and its movement determine the reaction of Neuro-Baby, including its facial expressions and actions (these facial expressions and actions were previously created by an intuitive design process developed by one of the authors). These reactions are visualized with computer graphics along with appropriate speech output.

(1) *Configuration of neural network.* The neural network for emotion recognition is a combination of eight sub-networks (Fig. 3). The decision logic stage combines the outputs of these sub-networks and outputs the final recognition result. Each sub-network is tuned to recognize one of seven emotions (anger, sadness, happiness, fear, surprise, disgust, and teas-

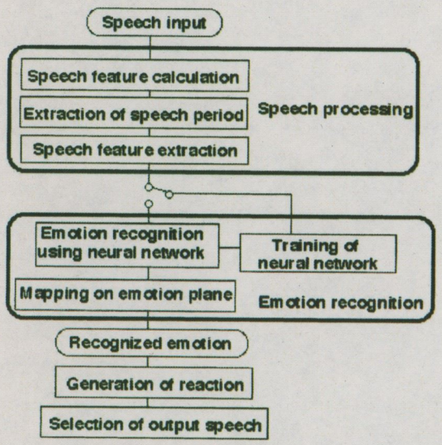


Fig. 2. Block diagram of processing flow.

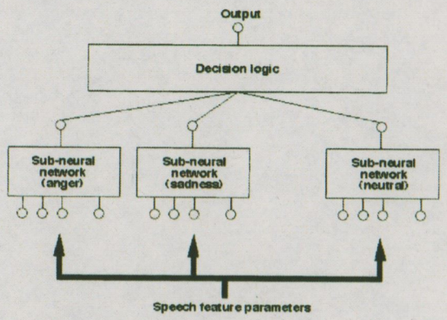


Fig. 3. Configuration of emotion recognition.

one else's mind (Murray, 1997). In order to develop empathy-inducing technology, the author first considered using an interactive computer character to achieve this state of mind. A number of recent studies have focused on interactive characters and agents (Bates, 1992; Maes et al., 1996; Perlin, 1995; Reilly, 1996), but few of them were designed to create empathy. In this study, an effort was made to develop a contextualized and creative system that enables us to enter the world of a movie to enjoy the atmosphere and talk with the characters and that also changes the situation in response to the communications.

Lifelike emotional character 'Neuro-Baby: MIC' (Tosa and Nakatsu, 1996)

Concept

In this fast-paced information society, you may have felt a longing for a 'partner' that, unlike a person or a pet, doesn't demand much attention or care but still has a 'mind' able to make its own 'judgment'. First I designed Neuro-Baby to provide such a partner, to act as your alter ego or the person closest to you. Neuro-Baby can judge emotions and respond to them (Tosa, 1993).

Neuro-Baby has advanced from its first version through several stages (Tosa et al., 1994) to 'MIC' (Fig. 1). In our present research, we tried to realize higher-level processing that could achieve more sophisticated interactions between Neuro-Baby and humans.

Fig. 1. People talking to Neuro-Baby: MIC.

Design of emotion model

How many and what kinds of emotional expressions to be adopted are both interesting and difficult issues. In our previous study, we investigated four emotional states (Tosa et al., 1994). Based on our experiences of demonstrating the first version Neuro-Baby to a variety of people and on the belief that more emotional states would make the interaction between Neuro-Baby and humans richer.

MIC recognizes the following emotions from intonations in the human voice. An asterisk indicates how the user should make intonations. The physical form of intonation is called prosody, and how to treat prosody will be discussed below.

(a) Joy (happiness, satisfaction, enjoyment, comfort, smile) * exciting, vigorous, voice rises at the end of a sentence

(b) Anger (rage, resentment, displeasure) * voice falls at the end of a sentence

(c) Surprise (astonishment, shock, confusion, amazement) * screaming, excited voice

(d) Sadness (sadness, tearfulness, sorrow, loneliness, emptiness) * weak, faint, empty voice

(e) Disgust * sullen, aversive, repulsing voice

(f) Teasing * light, insincere voice

(g) Fear * frightened, sharp, shrill voice

Communication

In most cases, the content of a media transmission conceals the actual functions of the medium. This content is impersonating a message, but the real message is a structural change that takes place in the deep recesses of human relations. We aim for this kind of deep communication. People use a microphone when communicating with MIC. For example, if the participant whistles, MIC's feeling is positive and he responds with excitement. If the speaker's voice is low and strong, MIC's feeling is bad and he gets angry.

Processing

This section describes the principles and operational details for the recognition of emotions included in speech. It also, explains the generation process of MIC reactions, which correspond to the emotions it receives.

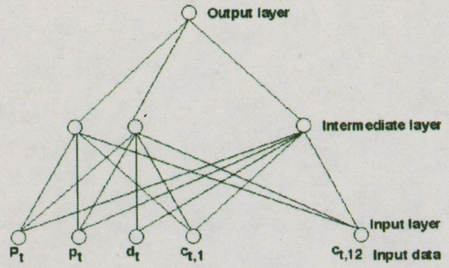


Fig. 4. Configuration of a sub-network.

ing) and neutral emotion. Basically, each sub-network has the same network architecture (Fig. 4), which is a three-layered neural network with 150 input nodes corresponding to the dimension of speech features, 20 to 30 intermediate nodes, and one output node.

(2) *Neural network training.* To recognize emotions, it is necessary to train each of the sub-networks. Since our target is speaker-independent and content-independent emotion recognition, the following utterances were prepared for training.

Words: 100 phoneme-balanced words
Speakers: five male speakers and five female speakers
Emotions: neutral, anger, sadness, happiness, fear, surprise, disgust, and teasing
Utterances: each speaker uttered 100 words eight times

In each of the eight trials, he/she uttered words using different emotional expressions. Thus, a total of 800 utterances for each speaker were obtained as training data.

Using these utterances, we carried out various preliminary-training tests. It turned out that preparing two kinds of networks for each emotion, one for male speakers and the other for female speakers, was better than preparing only one network to handle both male and female utterances. In other words, the emotional expressions between males and females are somewhat different and cannot be handled together. The reason for this is not clear and will require further research.

(3) *Emotion recognition by a neural network.* In the emotion recognition phase, speech feature parameters extracted in speech processing are simultaneously fed into the eight sub-networks and trained as described above. Eight values, $V = (v_1, v_2, \ldots, v_8)$, are obtained as the result of emotion recognition.

(4) *Mapping on an emotion plane.* As described above, the output of the emotion recognition network is a vector $V = (v_1, v_2, \ldots, v_8)$, and the final recognition result should be obtained based on V. In our previous study, we expressed the final emotion state by a point on a two-dimensional plane. Based on the experiences of previous research, in the present study the positions of the eight emotions were rearranged on emotion plane E as shown in

It is necessary, therefore, to carry out the mapping from V onto E. Let m_1 and m_2 be the first and second maximum values among $v_1, v_2, \ldots, v_8$, and also let (xm_1, ym_1), (xm_2, ym_2) be the emotion positions corresponding to m_1 and m_2, respectively. The final emotion position (x, y) is calculated by

$$x = c * xm_1 + (1 - c) * xm_2,$$
$$y = c * ym_1 + (1 - c) * ym_2$$

(c : constant value).

Reaction of the characters

The reactions of MIC could be carefully designed and were visualized with computer graphics. Several examples of emotional expressions by MIC are shown in Fig. 5.

Interactive Poem (Tosa and Nakatsu, 1997)

Concept

The multi-modal character MIC was designed to generate complex emotions but not to recognize meanings. But people look for meanings in conversations, so they inevitably talk to MIC, which can't understand their words. Even a child has the ability to make up meaningful 'conversations' with a doll or even something that is not lifelike. Accordingly, nonverbal communications alone can't offer precise and deep communications. Therefore, I tried to design a system for verbal communications with a computer character. It is extremely difficult to conduct the

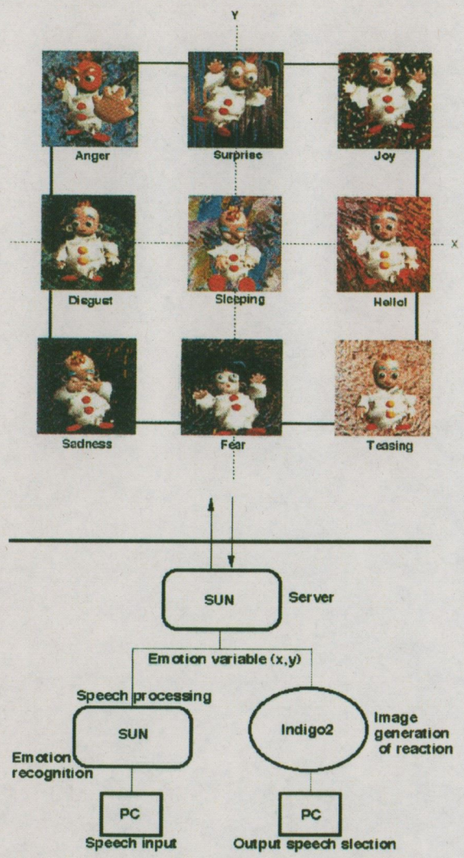


Fig. 5. Emotion expression by MIC.

emotional Japanese-style conversation based on deep mutual understanding in real time with a computer. When I was groping for a design of a framework that could serve as the basis for verbal communications with a computer character, the idea of using the traditional 'renka' linked poem formula struck me. The idea developed into the concept of 'interactive poems,' improvised poems composed alternately by a computer poet and a human using the renka formula. Poems are stylized verbal expressions of a poet's messages; we can enjoy the world according to the poet by reading his poems. It is assumed that interactive poems (Tosa and Nakatsu, 1998) jointly composed by a person and a computer will lead to unexpected happenings (accidentality) and show the differences in personality among the composers. Such results would result from the effects of interaction. Accordingly, a medium for communications based on sensibilities can be created by incorporating the functions of a 'dialogue.'

Interaction

The face of Muse, the Greek goddess of music, appears in a large screen. The Muse spins poems one after another while talking with a person in a singing manner. The Muse talks to the user in a very emotional way while using brief poetic phrases. Her words help the user enter her world and inspire him to respond with his own poetic words. Through this process of exchanging poetic words according to the formula of 'Interactive Poems,' the user and the computer can work together to create a world of improvised poems filled with inspirations, feelings and sensibilities (Fig. 6).

System configuration

The Interactive Poem system is comprised of four units, including (1) a system control unit, (2) a voice recognition unit, (3) an image generation unit, and (4) a voice output unit (Fig. 7).

The system control unit uses a database and controls the operation of the entire system. In this system, the most important feature is the construction of interactive poems.

First, an explanation will be given of the contents of the database, which houses the interactive poems. As a premise, it is believed that poems of the past are continuous poetic phrases. In other words, the basic construction of these poems can be expressed through a simple conditional transition network. With this conditional transition network, each phrase accommodates conditions, and each of those conditions connects to another condition (Fig. 8).

The basic forms of interactive poems are expressed through a transition network. The difference from a conventional transition network is that

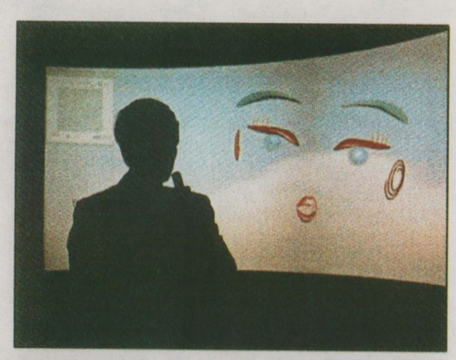


Fig. 6. People playing with interactive poem.

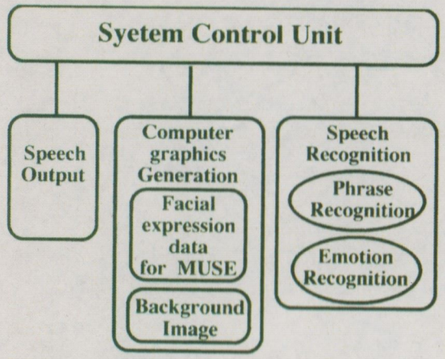
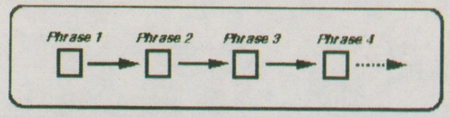


Fig. 7. Block diagram of interactive poem system.



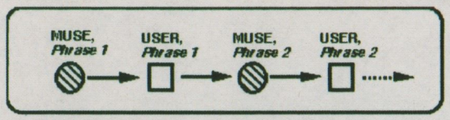
You are standing
You head for the sea
You gaze out
A small road is passing
A small road I have never walked
On this small road
A person I have never met
That person turns towards you
That evening
It is the light off an aluminum pot
On the top of kotatsu
that has been delivered
It is the tears
you love
I didn't see this
Even if you became
Even if you told me .....
.....

Fig. 8. Conventional poem.

phrases uttered by the computer and phrases uttered by the user are expressed alternately. The computer and user can alternately read the phrases of poems determined beforehand for a simple conversation (Fig. 9).

With this interactive poem system, we prepared many phrases that could be linked to phrases uttered by the computer to achieve a higher degree of conversation. These phrases were considered and made with the utmost care to have phonemic and meaningful relations with the computer phrases. This transition network is put in the database and used to control all processes. Through this construction, the user will select phrases meeting his or her own emotions and 'Japanese spiritual emotion.' By speaking to the computer, the user can add his or her own emotions and 'Japanese spiritual emotion' to the world of an original poem (Fig. 10).

The voice recognition unit has two functions: one for meaning recognition and the other for emotion



M: Who are you?
Not I but you
Not him but you

U: Someone other than I
M: with the same ears as mine

U: who hears
M: different sounds from mine

U: Someone who has ten fingers
M: just like mine

U: who tries to grasp
M: what I cannot

M: You

U: You're standing
M: in mid-summer sunlight

U: facing the sea
M: your back to me

U: You're looking
M: at the distant horizon

U: In your mind
M: runs a lane

U: I've never walked on
M: of a town I've never seen

U: Snow is now
M: quietly falling on that lane

U: and someone I've never met
M: is running towards me

U: What he yelled at you
M: I shall never know
M: never

*M:Muse U:User

Fig. 9. Construction of interactive poem (a).

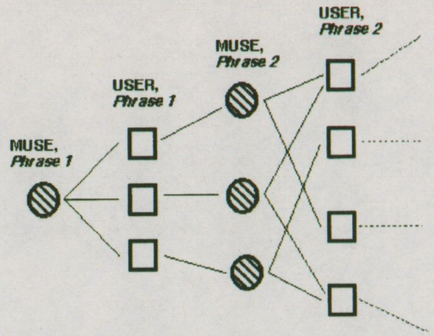


Fig. 10. Construction of interactive poem (b).

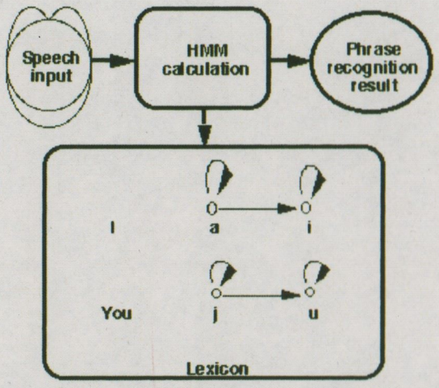


Fig. 11. Voice recognition system.

recognition. To recognize the meanings of phrases uttered by a user, the unit uses non-specific talker voice recognition based on a Hidden Markov Model (HMM) (Shimizu et al., 1996). Each phrase that is uttered is displayed in the form of a series of phonemes and recorded into a recognition dictionary (Fig. 11).

At the same time, the unit performs emotion recognition to recognize the emotion information of the user. The emotion recognition technology used here is the same with the one used for MIC (Fig. 12).

Then, the computer poet MUSE's generated emotional expression depends on the emotional states recognized by the neural network (Fig. 13). This neural network is the same as that used for Neuro-Baby 'MIC'.

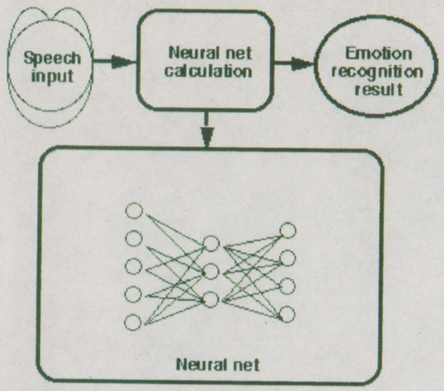


Fig. 12. Emotion recognition system.

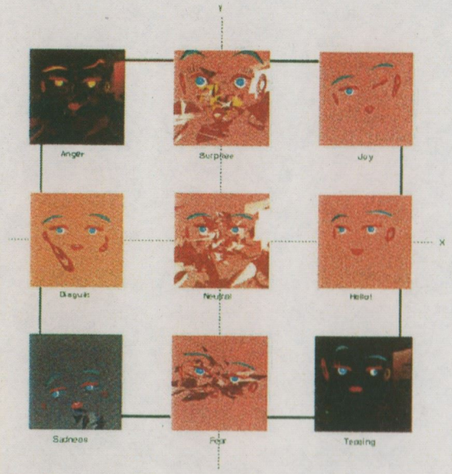


Fig. 13. Reaction of emotional expression of MUSE.

Emotion recognition-based, Multi-person Interactive Theater System (Nakatsu and Tosa, 1999)

Concept

The author has met the challenge of communications involving a number of humans and a computer by developing interactive movie-specific methods, not stopping at the one-to-one communications of a human and a computer. From a different perspective than conventional views, work was done on future movie media research and example content production (Jacob, 1997). Concretely speaking, conventional movies take the form of giving an imaginary world (i.e., a cyberspace) and a story to an audience in a unidirectional manner. By adopting interaction technologies, in contrast, viewers themselves can become the main characters, can enter the cyberspace, and can experience the stories autonomously.

Fig. 14. People playing *Romeo and Juliet* in International Berlin Film Festival.

From this viewpoint, the author will explain the structure of an interactive movie system that introduces interaction technologies to conventional movies and a sample of artwork as an actual example (Fig. 17).

We selected *Romeo and Juliet* by Shakespeare as the base story. The main plot of the story is as follows. After their tragic suicide, the souls of Romeo and Juliet are sent to Hades, where they have no recollection of the past. Then, Romeo and Juliet each start on a journey to rediscover themselves and what relationship they shared with their characters, who are emotion-based autonomous actors (Fig. 14).

Software that realizes conscious/unconscious interactions

Participation in cyberspace

(a) System for multiple players: Our initial efforts in making a system for multiple players allowed two

Fig. 15. People controlling avatar by motion capture system.

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. However, the first step for the present study is to develop a prototype multi-player system that has two systems connected by a LAN (Fig. 18).

(b) Avatar representation: We used a system that showed avatars on screen as alter egos of the players (Fig. 15). There were several advantages to this system, as outlined below.

(b-1) System for representing the avatar: The relationship between the player and his 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.

(b-2) System for controlling the avatar: The basic control system inputs player movement by using magnetic sensors and uses that movement to map avatar movement with a motion capture system (Fig. 15). Giving autonomy to avatar movement enables a complex movement that combines autonomous avatar movement with player movement. By varying the proportion of each movement based on time and circumstances, player movements can be used directly. In addition, the player can introduce desired movements that add diversity and depth to the relationship with cyberspace.

Interaction

(a) 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 the characters and generally do not affect story development. On the other hand, there are sometimes interactions that can affect the story development. These kinds of interactions, called story conscious interactions (SCI), occur at branch points in the story. The results of such interactions determine the outcome of the story.

(b) Introduction of multimodal interaction: The following interactive functions were added to the primary interactive function of voice recognition.

(b-1) Emotion recognition: To make interaction possible at any time, an emotion recognition capability was introduced. When players utter spontaneous utterances, the characters in the story react with their own utterances and animation according to the emotion recognition results. Emotion recognition is achieved by using a neural-network-based algorithm (Tosa and Nakatsu, 1996). People use a headset microphone when communicating with characters. For example, if one makes a happy high voice, the character's feeling will be positive and he responds with excitement. If the speaker's voice is low and strong, the character's feeling will be bad and he gets angry.

(b-2) Motion capture: We used a motion capture system based on magnetic sensors attached to appropriate parts of the player's body. This system allows avatar movement to reflect player movement. Data from the magnetic sensors are input into the system to move the computer graphic avatar. In this way, players get the feeling that they are controlling the movements of their avatar. This control can provide another form of interaction at any time.

(b-3) Gesture recognition: We captured motion from the players with magnetic sensors and used a HMM to process the data from the sensors. This was done so that 3-D gestures, minute gestures,

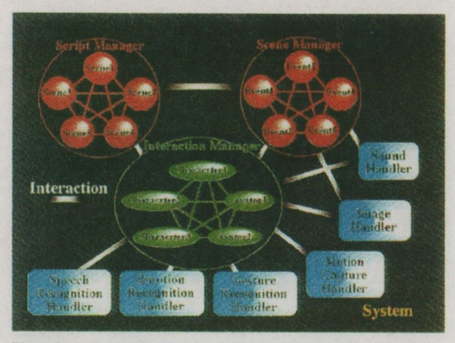


Fig. 16. Authoring system for Interactive Theater.

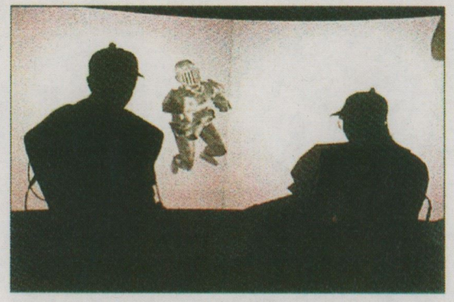


Fig. 17. One of the scenes of Interactive Theater.

and gestures under low-light conditions could be recognized. Gesture recognition results are used for SCI. Fig. 16 shows an overview of a system equipped with these functions.

Generation of empathy by interactive story

The scenario in this study is devised by the people playing Romeo and Juliet, who select lines depending on their mood. How the drama unfolds depends on this selection process. If the people playing the roles of Romeo and Juliet speak words of anger or of happiness, depending on the context of the play, autonomous characters recognize the emotions from the tone of voice and react with their emotions through speech and animation. Emotional recognition technology allows a neural network to learn emotional speech, and through this learning, the network can create personalities such as an angry character or a cheerful character.

Voice recognition is the function whereby lines are chosen and said by a player to go with the scene and then the autonomous character recognizes the meaning of these lines and reacts to them with utterances and gestures (Tosa and Nakatsu, 1998).

Gesture recognition is the reaction of the character to action, contact or a pose in any location. For example, if the impassioned Romeo tried to kill his former friend Mercutio with a pistol, Mercutio would recognize this behavior, escape and condemn Romeo.

- Follow interaction: This function allows the characters in the movie system to follow or accompany the avatar of a player in an appropriate situation. For example, if the player who plays the role of heroine Juliet is in a serious situation, the autonomous agent Shin, an angel, recognizes her mental situation by emotion recognition, joins her, and takes care of her.
- Background speech interaction: This function allows you to talk from the background to a character about your feelings and doubts concerning the performance running in the foreground.
- Touch interaction: This function is used when, for example, the lead actor Romeo becomes very angry and hits his old enemy Paris. Paris then pulls back in fear. Fig. 17 shows one of these scenes. In this scene, Paris, who was once Juliet's fiancee is being killed by his enemy. Romeo and Juliet try to help him using voice and gesture recognition.

Hardware system structure

Fig. 18 shows the system's hardware structure, which is composed of subsystems for image output, voice and emotion recognition, gesture recognition and sound output. Two workstations (Onyx Infinite Reality and Indigo 2 Impact) capable of generating computer graphics at a high speed are used to output the images. The Onyx is used to run the script manager, scene manager, interaction manager, and all of the image output software. The character images are stored on the workstations ahead of time in the form of computer graphic animation data so that computer graphics can be generated in real time. The

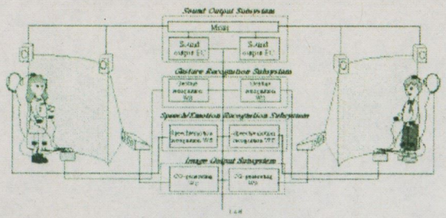


Fig. 18. Appearance of multi-player system.

background computer graphic images are also stored as digital data so that background images can also be generated in real time. The background images are real photographic images 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 eoscopic vision control and projected onto a curved screen with two projectors.

Result

Our system/work, *Romeo and Juliet in Hades* was selected for exhibition at Transmediale'99, which is a new media division of the International Berlin Film Festival 1999 (Fig. 14). We have developed both Japanese and English versions, demonstrated the English version at the exhibition. More than five hundred people from various countries visited our exhibit, and about two hundred of them played Romeo or Juliet. We requested these users to answer a questionnaire that asked the following.

1. How did you feel while playing Romeo or Juliet?
2. What should be added to the system for you to experience deeper empathy?
3. Other questions.

For the first question, almost all people said that the experience was new and interesting. This indicated that Romeo and Juliet in Hades was successful.

For the second question certain, a number of people said that they sometimes felt some difficulty in controlling their avatar and following the story.

Users also indicated that it is necessary to introduce an easier human interface. In our system, because people are both observers and players, they are a little bit busy for rather preoccupied with watching the ongoing story and playing their roles at the same time. At the same time, young people felt no difficulty because they are used to the same kind of experience when playing video games. If we want this system used both by young and older people, further research will be necessary to introduce a more human-like interface.

In other responses, some people wanted more touching feeling and pain feeling in the story. Other people wanted more communication between Romeo and Juliet.

Unconscious flow (Tosa, 1999)

Concept

In face-to-face communication, the occasional need for intentional lies is something with which everyone can identify. For example, when we get mad, circumstances may force us to put on a big smile instead of expressing our anger (Fig. 19); when we feel miserable, good manners may dictate that we greet others warmly. In short, to abide by social norms, we consciously lie. On the other hand, if we consider the signs that our bodies express as communication (body language), we can say that the body does not lie even when the mind does. Considering this phenomenon, we propose a means of 'touching the heart' in a somewhat Japanese way by measuring the heartbeat of the 'honest' body and using other technologies to develop a new code of non-verbal communication from a hidden emotion in society.

Fig. 19. Hidden emotion of people.

Fig. 20. In the bucket of "Unconscious Flow".

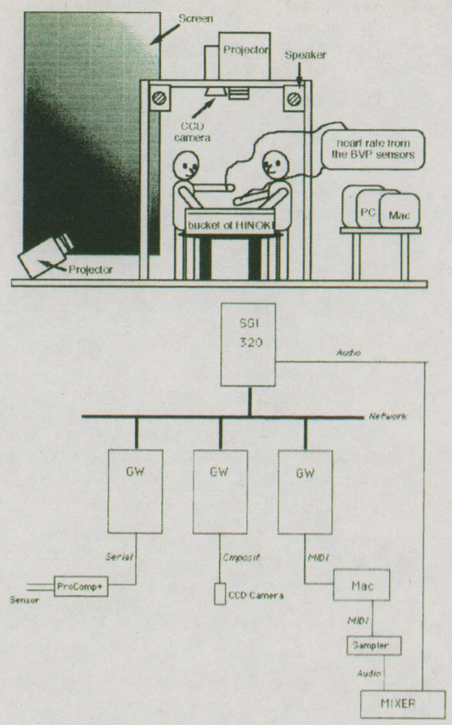


Fig. 21. System of unconscious flow.

Interaction

Two computer-generated mermaids function as individual agents for two viewers. Each mermaid agent moves in sync with the heart rate detected by an electrode attached to the collarbone of its viewer. Then, using a synchronization interaction model that calculates the mutual heart rate on a personal computer, the two mermaids express hidden non-verbal communication. The data of relax-strain calculated from the heart rate and the interest calculated from the variation in the heart rate is mapped on the model. The synchronization interaction model reveals the communication codes in the hidden dimensions that do not appear in our superficial communication. Then, using a camera to pick up hand gestures and a personal computer to analyze the images, the synchronization interaction model is applied to determine the mermaid's behavior. For a high degree of synchronism, the agents mimic the hand gestures of their subjects, but for a low degree of synchronism, the agents run away (Fig. 20). In the event that one mermaid agent touches the other, a pseudo-touch can be felt through the use of a vibration device. As for background sound, the heart sound of the subjects are picked up by an electronic stethoscope and processed for output on a personal computer. This work was done in collaboration with SONY-Kihara Research Center Inc.

System

For installation (Fig. 21), a space of four meters wide, four meters deep and three meters high is required. A dark and quiet space is preferable. Interactive actions are displayed on one main screen and two Japanese 'shoji' screens. A Japanese 'hinoki' wooden bucket with a diameter of one meter is filled with water and placed in the center of the installation. Two persons, fitted with a stethoscope, experience non-verbal communication by touching their Compeer Graphics embodiments in the bucket. The PC calculates the synchronicity based on the heart rate from the electrodes of the electrocardiograph, and the PC generates an arbitrary feeling in a Compeer Graphics form. The hand movements of the two persons are captured by an installed camera, and an image analysis of the data is performed. In accordance with the synchronicity interaction model, the Compeer Graphics embodiment either follows the

Fig. 22. Unconscious flow at SIGGRAPH'99.

movement of the hand of the partner with high synchronicity or goes away from the hand of the partner with low synchronicity. When one touches the Compeer Graphics embodiment of the partner, a vibrator gives the partner a simulated feeling of touch. The heart rate sensor measures the timing of the heart, which is processed by the PC and outputted.

This system was exhibited at SIGGRAPH'99 Art Show (Fig. 22). Many people of diverse nationalities interacted with Unconscious Flow.

Synchronization interaction model

The data of relax-strain calculated from the heart rate and the interest calculated from the variation of the heart rate are mapped on the model. The synchronicity interaction model reveals the communication codes in the hidden dimension that do not appear in our superficial communication (Fig. 23).

For example, (1) when both people are in the domain where they are highly relaxed and interested, they are considered synchronized. Animation is generated in which, for example, their Compeer Graphics-reactive embodiments join hands in companionship or enjoy friendly actions (Fig. 24).

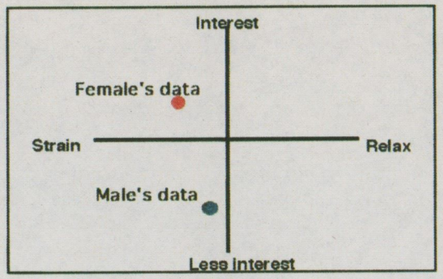


Fig. 23. Synchronicity interaction model.

(2) When both people are in a situation where they

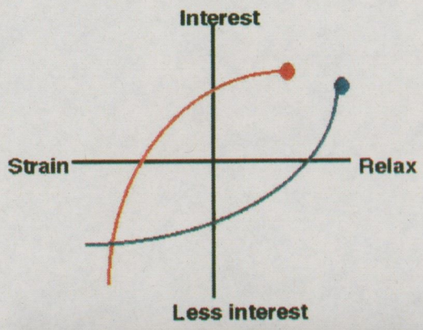


Fig. 24. Highly strained and highly interested.

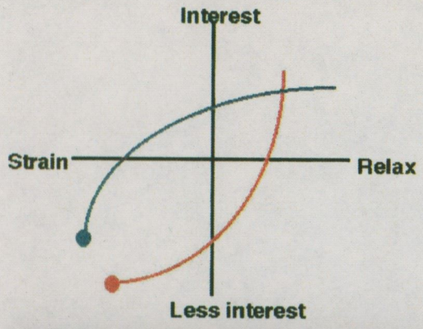


Fig. 25. Highly strained and less interested.

are highly strained and less interested, unfriendly communication is generated. Animation is generated in which, for example, their Compeer Graphics embodiments quarrel with each other (Fig. 25).

Hand recognition

A CCD camera recognizes each person's hand by a marker. The Image Recognition program processes the distance between the two markers' positions detected by the camera and determines whether a hand touches a mermaid or not (Fig. 26).

Heart rate sensor

A person's heart rate is measured by placing the Heart Rate Sensor (Fig. 27) on his finger. The heart rate is sent to a PC connected to the sensor (Pro-Comp+) via RS232C and is mapped on the synchronicity model.

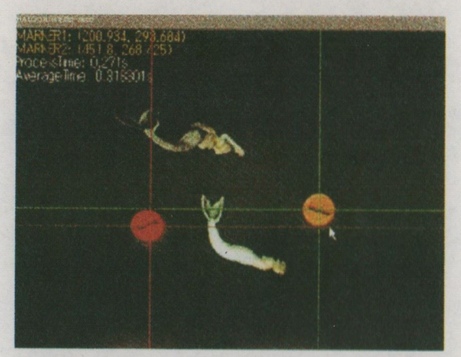


Fig. 26. Image processing by hand detection.

Fig. 27. People wearing heart rate sensor.

Software configuration (Fig. 28)

The Heart Rate Analyzer (Fig. 27) analyzes the input data and sends it to Event Control as event data. Event Control sends heart rate data as MIDI commands to the MAX/MSP program on a Macintosh and some commands to the Computer Graphics Generator if some computer graphics need to be changed depending on the heart rate. Computer Graphics Generator creates computer graphics based on these commands and then outputs the computer graphics. The MAX/MSP program processes the sound data and the heart rate sound as required and then outputs the resulting sounds. Image Recognition analyzes the image data fed from the CCD camera and the

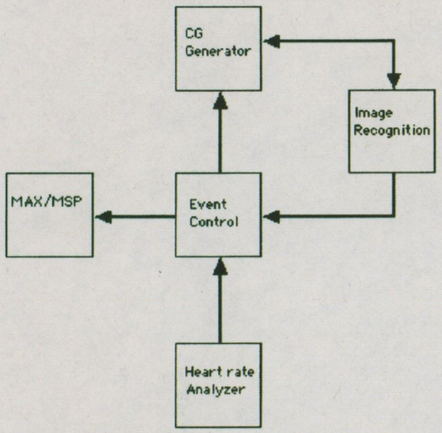


Fig. 28. Software configuration.

relational information of the hand, and the computer graphics displayed are sent to Event Control and Computer Graphics Generator. Event Control again sends some commands to Computer Graphics Generator if some computer graphics need to be changed depending on the data.

Evaluation

One of the achievements of this study was the creation of a human–computer interface through emotions. As mentioned in the section **Lifelike emotional character 'Neuro-Baby: MIC'**, I first created 'Neuro-Baby MIC', a lifelike computer character that recognizes human emotions and responds to them. This capability of MIC represents the first example of an emotion interface. 'Neuro-Baby MIC' has been demonstrated at a number of technology and art exhibitions both in Japan and abroad, receiving great praise from the people who tried the system. This showed that an emotion interface could work well regardless of the cultural environment.

This study has also succeeded in inviting human empathy with the computer character. The traditional human–computer interface is evaluated by efficiency. As human communications make clear, however, efficiency is not a good criterion for evaluating an emotion interface. In order to evaluate an emotion interface properly, one must see how well it can make people engrossed in interaction, emotionally involved, and feel deep empathy with the computer character.

The author created Neuro-Baby believing that turning emotions into an interface holds the key to stimulating empathy. As I had expected, or rather far beyond my expectation, people became engrossed in dialogues with Neuro-Baby and empathized with it. This shows that the study was on the right track from the beginning. Many of the people who had dialogues with the baby at one of the exhibitions said they had enjoyed very sophisticated interaction or that they had felt friendly affection toward the baby, which they said had human feelings.

For the next step, I tried to add verbal elements to the system. Since Neuro-Baby MIC can only offer non-verbal interaction, I tried to realize interaction based on understanding of words. The effort was focused on empathy. I chose the poem as the mode of communications because it particularly concerns emotions and sensibilities. From this viewpoint, I created a system for 'interactive poem' that enables joint composition of poems by a person and a computer using the traditional Japanese 'renga' formula. The system turned out to be a far greater success than I had expected. It was demonstrated at a host of art and technology exhibitions to study people's reactions. Most of the people who tried the system talked very positively about their experience, making remarks like "I completely forgot myself in exchanging poems with the Muse" or "I could enter much deeper into the spirit of poems than when I am only reading them silently." Particularly notable was the fact that the English version of Interactive Poem, with an English translation of the original poem composed by a Japanese poet, had a huge emotional impact on American and European audiences at overseas exhibitions. This suggests that an effective synthesis of verbal and non-verbal interfaces can induce deep empathy in the people exposed to it irrespective of their cultural backgrounds.

In Interactive Theater, I introduced a drama to create even deeper empathy. It is well known that when people read a novel or watch a film they deeply empathize with the characters in them. But we can never really enter the worlds created in novels

or movies, however strong our empathy may be, because of the solid wall that separates a real world from a virtual one. In this study, the introduction of interaction helped break down the wall and create an enchanting sense of real presence within the virtual world. I created a contextualized and creative system that allows people to enter the world of the movie, taste the atmosphere, and talk with the emotion characters while changing the drama in response to their talks. Interactive Theater is designed to let people get immersed in a virtual world and have 'first-hand' experiences in this world.

Interactive Theater has been shown at many film festivals and exhibitions at the invitation of the organizers, including the Berlinale, the prestigious film festival, and Imagina2000, the largest computer graphics conference in Europe. The audience responses were enthusiastic. A majority of the people talked about their impressions in language indicative of deep emotional experience. "I was moved when an emotion character responded immediately to my lines." "It has succeeded in creating narrative cyberspace." "I was thrilled and excited." "I was raring to speak my part." Their words have convinced me that Interactive Cinema induced the audience's empathy with the characters in cyberspace in a way that was much deeper than had ever been achieved before.

Unconscious Flow was exhibited at SIGGRAPH'99, which was held in Los Angeles. A large number of people visited the exhibition site and enjoyed the interaction of Unconscious Flow. On the West Coast, the idea of healing and meditation is familiar, which I believe explains why this work was positively accepted by so many people. So far, this work uses a biofeedback function based on one's heart rate. Other areas related to this work yet to be explored include sound hearing and hearing psychology.

Conclusions

Emotion information has properties like subjectivity, polysemy, vagueness, and dependence on the situation (Heilig, 1992). Being subjective means it depends on individuals, raising the question of preferences. Emotion information can have many different meanings (polysemy), depending on the receivers. It is also greatly affected by the environment, which means its meanings depend on the situation. Since emotional responses contain information about the subconscious mind, it is important to seek the deeper causes of emotions. That's why emotion information cannot be symbolized without undermining its significance. Its assessment should be based on criteria that take account of individual and cultural differences. Specifically, research is needed on a natural input method responding to the causes of emotions, the construction of a database composed of emotion information, and the effects of various situations on the formation of specific emotions.

The author's expression consists in the generation of unknown imagination that lies between man and the computer. The experience doesn't necessarily have a happy ending. An automated character may betray the man. The man himself may gripe about his problems or feel horror or sorrow. But the experience always entails a sense of fulfillment and catharsis. That will eventually lead to research into unknown aspects of mankind. It is said the machine that comes closest to a living form is the most beautiful. A Compeer Graphics character can move people when it has a human touch. One of the major emotions a man feels when he gets a flash of intuition is aesthetic satisfaction. Beauty is truth, and truth is beauty. Senses of beauty and harmony put vibrant life into art and science. This is an era when science needs sensitivity to beauty.

Research on the technology to enable computers to explore and grasp human sensibilities, emotions and the depths of man's consciousness will advance and expand in coming years. But man's mind and body have aspects that defy any generalization and formalization. A simple technological approach is no longer enough for research on human perception or expansion of sensibilities. The history of art is the history of spiritual and sensory expressions of human perception and the expansion of sensibilities. Excellent works of art are often created with leading-edge tools. When we are moved by a novel or a movie, our mind is exploring in a virtual world. There is no engineering technology available yet to induce empathy. That's because any such system would be based on the private views of the creator and involve spiritual and philosophical problems. Effort should be devoted to engineering the design of such

a system. Organizing sensibilities into an effective system is more important for enhancing originality than research on computer analysis of sensibilities. The development of multimedia has made it possible to turn methodologies of artistic expressions into engineering processes by using computers. People who have never mastered a method of expression can now express images with a computer. The technology can make a great contribution to mankind as an effective and high-potential medium for communicating images to others. It will have such a huge impact on the social system as to change its structure and also change the nature of links and relations in the present order of things and the form of information in society.

Acknowledgements

Neuro-Baby 'MIC' and Interactive Poem and Interactive Theater were developed in collaboration with Dr. Nakatsu from ATR Media Integration and Communication Research Labs. Unconscious Follow was done in collaboration with Sony-Kihara Research Center, Inc. Special thanks are given to manager Mr. Ueda and senior researcher Mr. Asukai, Mr. Sakamoto, Mr. Ohto, Mr. Nozaki, Mr. Serita, and general manager Mr. Komatsu. In addition, ambient heart beat music was provided by Mr. Nagahara and real-time computer graphics technology was provided by Mr. Ozaki of Sony Creative Center.

References

Abadjieva, E., Murray, I.R. and Arnott, J.L., Applying analysis of human emotional speech to enhance synthetic speech. In *Proceedings of IEEE*, Berlin, Germany, 1993, pp. 909–912.

Bates, J., Virtual reality, art, and entertainment. *Presence: J. Teleoperators Virtual Environments*, 1 (1992) 133–138.

Bates, J. et al., An architecture for action, emotion, and social behavior. In *Proceedings of the Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World*, 1992a.

Bates, J., Loyall, A. and Scott Reilly, W., Integrating reactivity, goals and emotions in a broad agent. Technical Report CMU-CS-92-142, School of Computer Science, Carnegie Mellon University, Pittsburg, PA, May 1992b. Also appeared in *Proceedings of the 14th Annual Conference of the Cognitive Science Society*, Bloomington, IN, July 1992b.

Brooks, R.A., From earwigs to humans. *Proceedings IIAS, the 3rd Brain and Mind International Symposium "Concept Formation, Thinking and their Development"*, Kyoto, Japan, May 1996, pp. 59–66.

Brooks, R.A., Breazeal, C., Marjanovic, M., Scassellati, B., Williamson, M., The Cog Project: building a humanoid robot, *Springer Lecture Notes in Computer Science*, 2000, in press.

Hayes-Roth, B. and van Gent, R., Story-making with improvisational puppets and actors, *Stanford Knowledge Systems Laboratory Report* KSL-96-05, 1996.

Heilig, M.L., The cinema of the future, *Presence*, 1 (1992) 279–294.

Isbister, K. and Hayes-Roth, B., Social implications of using synthetic characters: an examination of a role-specific intelligent agent. Stanford Knowledge Systems Laboratory Report, KSL-98-01, 1998.

Jacob, O., Computer graphics story — a personal overview of computer animation in the movies, *ACM Comput. Graph.*, 31 (1997) 26–28.

Kobayashi, H. and Hara, F., Analysis of the neural network recognition characteristics of 6 basic facial expression. In *3rd IEEE International Workshop on Robot and Human Communication*, 1994, pp. 222–227.

Laurel, B., *Computer as Theater*, New York, Addison Wesley, 1992.

Leopoldseder, H. and Schopf, C., Cyber arts. *Prix Ars Electronica Edition 97*, Berlin, Springer, 1997.

Maes, P., Darrell, T., Blumberg, B. and Pentland, S., The ALIVE system: full-body interaction with autonomous agents. In *Proceedings of the Computer Animation 95 Conference*, Geneva, Switzerland. IEEE Press, April 1995, pp. 11–18.

Maes, P., Darrell, T., Blumberg, B. and Pentland, A., The ALIVE system: wireless, full-body interaction with autonomous agents. To be published in a Special Issue on Multimedia and Multisensory Virtual Worlds, ACM Multimedia Systems, ACM Press, Spring 1996.

Morishima, S. and Harashima, H., Emotion space for analysis of facial expression. In *IEEE International Workshop on Robot and Human Communication*, 1993, pp. 188–193.

Murray, J.H., *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*, New York, Simon and Schuster, 1997.

Nagao, K. and Takeuchi, A., Speech dialogue with facial displays: multimodal human–computer conversation. In *Proceedings of 32nd Annual Meeting of Association for Computing. Linguistics, (ACL-94)*. Morgan Kaufman Publishers, 1994, pp. 102–109.

Nakatsu, R. and Tosa, N., *Interactive Movies, Handbook of Multimedia Computing*. CRC Press, 1999, pp. 701–712.

Perlin, K., Real time responsive animation with personality, *IEEE Trans. Visual. Comput. Graph.*, 1 (1995) 5–15.

Picard, R.W., *Affective Computing*, MIT Press, Cambridge, MA, 1997.

Reilly, S., Ph.D. Thesis, Technical Report CMU-CS-96-138, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, 1996.

Reeves and Nass, *The Media Equation*, Cambridge, Cambridge University Press, 1996.

Scott Reilly, W. and Bates, J., Natural negotiation for believable agents. Technical Report CMU-CS-95-164, School of Computer Science, Carnegie Mellon University, Pittsburg, PA, June 1995.

Shimizu, T. et al., Spontaneous dialogue speech recognition using cross-word context constrained word graph. In *Proceedings of ICASSP 96*, 1996, 145–148.

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*, 1997, pp. 167–175.

Tosa, N., Talking to NEURO BABY. *ARS Electoronica'93 Genetic Art–Artificial Life Proceeding*, 1993, pp. 353–356.

Tosa, N. Unconscious flow. In *SIGGRAPH'1999 Electronic Art and Animation Catalog*, 1999, p. 11.

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*, 1996, pp. 12–19.

Tosa, N. and Nakatsu, R., Interactive poem. In *Proceedings of the AIMI Inernational Workshop on Kansei — The Technology of Emotion*, 1997, pp. 54–59

Tosa, N. and Nakatsu, R., Interactive poem system. In *The 6th ACM International Multimedia Conference Proceedings*, 1998, pp. 115–118.

Tosa, N. et al., Neuro-Character. In *AAAI '94 Workshop, AI and A-Life and Entertainment*, 1994.

Waters, K., A muscle model for animating three dimensional facial expression, *Comput. Graph.*, 21 (1987) 17–24.