ATR MIC JAPANESE

NAOKO TOSA

Contact

ATR Media Integration & Communication Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-02, JAPAN.
PHONE: +81 774 95 1427, FAX: +81 774 95 1408
E-MAIL:tosa@mic.atr.co.jp

LINK TO JAPANESE PAGES

Naoko Tosa is a Media Artist & Researcher in the ATR Media Integration & Communications Research Laboratories. She is also a guest professor in the graduate school of Science and Technology at Kobe University and a visiting lecturer at Musashino Art University's Department of Media Art & Science. She received Ph.D. of engineering for Art and Technology research from University of Tokyo. In particular, she focused on the topic of communication and used computers, video and electronics to design art work that relates to the intelligence of emotions, consciousness and unconsciousness. She specializes in the creation of experimental film, video art, computer graphics animation, and interactive arts. Her recent work includes the Neuro-Baby project, an autonomous computer agent with automatic facial expression and behavior synthesis that can respond to the human voice through recognizing emotions and feelings. Her work has been exhibited at the Museum of Modern Art New York, the New York Metropolitan Art Museum, SIGGRAPH, Ars ELECTRONICA, the Long Beach Museum, and other locations worldwide. Her works also part of the collections at the Japan Foundation, the American Film Association, the Japan Film Culture Center, the Nagoya Prefecture Modern Art Museum, the International Art Museum of Japan, the O Art Museum, the Toyama Prefectural Museum of Modern Art, and the Takamatsu City Art Museum. In 1996, she received the best paper award from the IEEE International Conference on Multimedia. In 1997, the L'Oreal Grand Prix for research combining art and science award her first prize.

Showing New Work

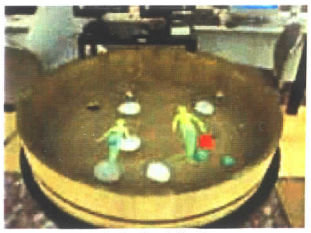
"Unconscious Flow"
1999
collaborated with Sony-Kihara Research Center, Inc.

SIGGRAPH 99 ART Show
Aug. 8-13 1999
Los Angels Convention Center
Los Angels, California USA

News

アートとテクノロジー国際会議

【“デジタル・ルネッサンス in けいはんな”】



土佐尚子作品
「無意識の流れ」

ペットロボット猫「たま」

アザラシ「ごま」
柴田宗徳（工業技術院機械技術研究所）

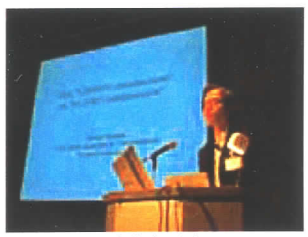
クリスタ＆ミニョーノ作品
「ヒコ・スキャン」

イマジナ'99作品
「ビンゴ」

国内に数ある学術研究都市の中で、唯一文化を名前に冠するけいはんな学研都市において、アートとテクノロジーの融合をテーマにした国際会議“デジタル・ルネッサンス in けいはんな”が、11月10、11日の両日、“けいはんなプラザ”を会場にして開催された。(財)関西文化学術研究都市推進機構と(社)関西経済連合会が主催し、(株)ATR知能映像通信研究所との共催によって実現した。企画／総合ディレクターをATR土佐尚子客員研究員が務め、この分野の世界の第一人者の工学系、アート系の講師を招聘し、インタラクティブアート作品とペットロボットの展示を行った。マサチューセッツ工科大学高等視聴覚研究所(MIT's Center for Advanced Visual Studies=CAVS)のステファン・A・ベントン所長による講演はCAVSの位置に関するもので、科学、技術、芸術、製造に加えてデザインといった分野の掛け橋となるCAVSの役目について話した。続いて、ATR土佐尚子客員研究員が、“先端技術で甦る共振感覚”と題した講演を行ない、アートとテクノロジーの具体的な融合事例として、感情を認識し表現する作品を、観客の分身が深層心理を表現する作品について紹介した。MIT人工知能研究所のロドニー・ブルックス所長による講演では、“人間とロボットの相互作用と未来のロボットの社会における地位”というタイトルからも分かるように、ロボットを人間とのインタラクションによって成長させるというところがユニークな点である。ATR知能映像研究所社長である中津良平氏による講演では、“感性コミュニケーションから悟りコミュニケーションへ”が行なわれた。ヨーロッパ最大の規模を誇るシーグラフと並んで有名なデジタル映像の国際会議『イマジナ』を主催するINA国立視聴覚技術研究所のジルベール・デュテルトル氏は、来年で20回目を迎える“イマジナ2000”について紹介した。１１日には、サイバネティクス、テレマティクス、インタラクティブメディアを芸術の分野に活用した草分け的存在の英国ウエールズ大学のロイ・アスコット教授の講演を始め、ATRクリスタ、ソムラー客員研究員は、人工生命と芸術作品におけるインタラクションと進化についての研究を講演し、事例として作品を展示した。ニューヨーク近代美術館の主任学芸員であるバーバラ・ロンドン氏は、ビデオとインタラクティブ・アートは今後の方向性や、電子作品収集の可能性について論議された。東京大学先端科学技術研究センターの廣瀬通孝教授は、”バーチャルリアリティ・テクノロジーとコンテンツ”という講演で、技術者にもクリエイティビティが重要であるということを強調された。最後にこの分野で約５０年間の経験がある岐阜県立国際情報芸術アカデミー学長の坂根巌夫氏により２０世紀におけるアート＆テクノロジーの歴史という観点から総括的講演が行われた。聴衆は、定員２００名のところ、４００名という多数の参加となり、その内訳は、工学系６割、芸術系２割、一般２割という状況であり、工学系の関心の高さが伺われた。

会場風景

ステファン・ベントン先生
(MIT)



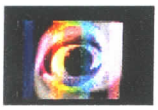
中津良平氏
（知能映像研究所社長）

ロドニー・ブルックス先生
(MIT)

ART WORKS

Video Arts

"TRIP"
8min. 1985



"Pleasure"
3min. 1986

"An Expression"
9min. 1985

"ECSTASY"
6min. 1987

"GUSH!"
6min. 1989

Interactive Arts

"Neuro-Baby"
1992-1994

Network Neuro-Baby"
995

"Neuro-Kids MIC"
1995

"Interactive Poem"
1996-1998

Interactive Theater
"Romeo & Juliet in Hades"
1998-1999

CAREER

CAREER

AWARDS

Referred Prize Winning History (-1989)
(1990-) Under Construction

PAPERS

International
Domestic (Japanese)

ATR MIC LABS' HOME
NAOKO TOSA tosa@mic.atr.co.jp

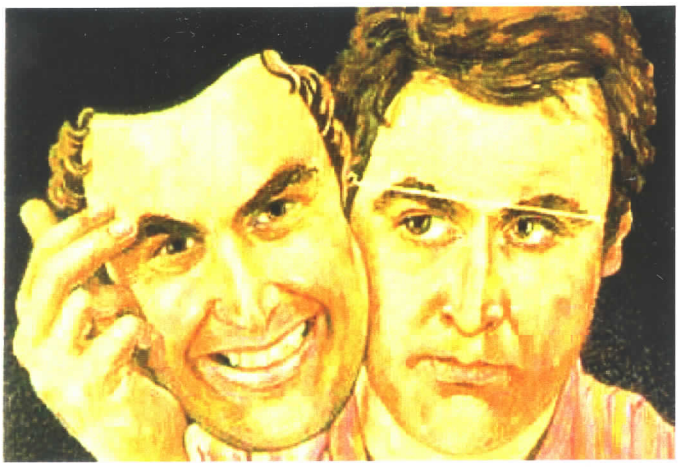
SIGGRAPH'99 ART Show

Title: Unconscious Flow

Artist: Naoko Tosa in collaborated with Sony-Kihara Research Center, Inc.

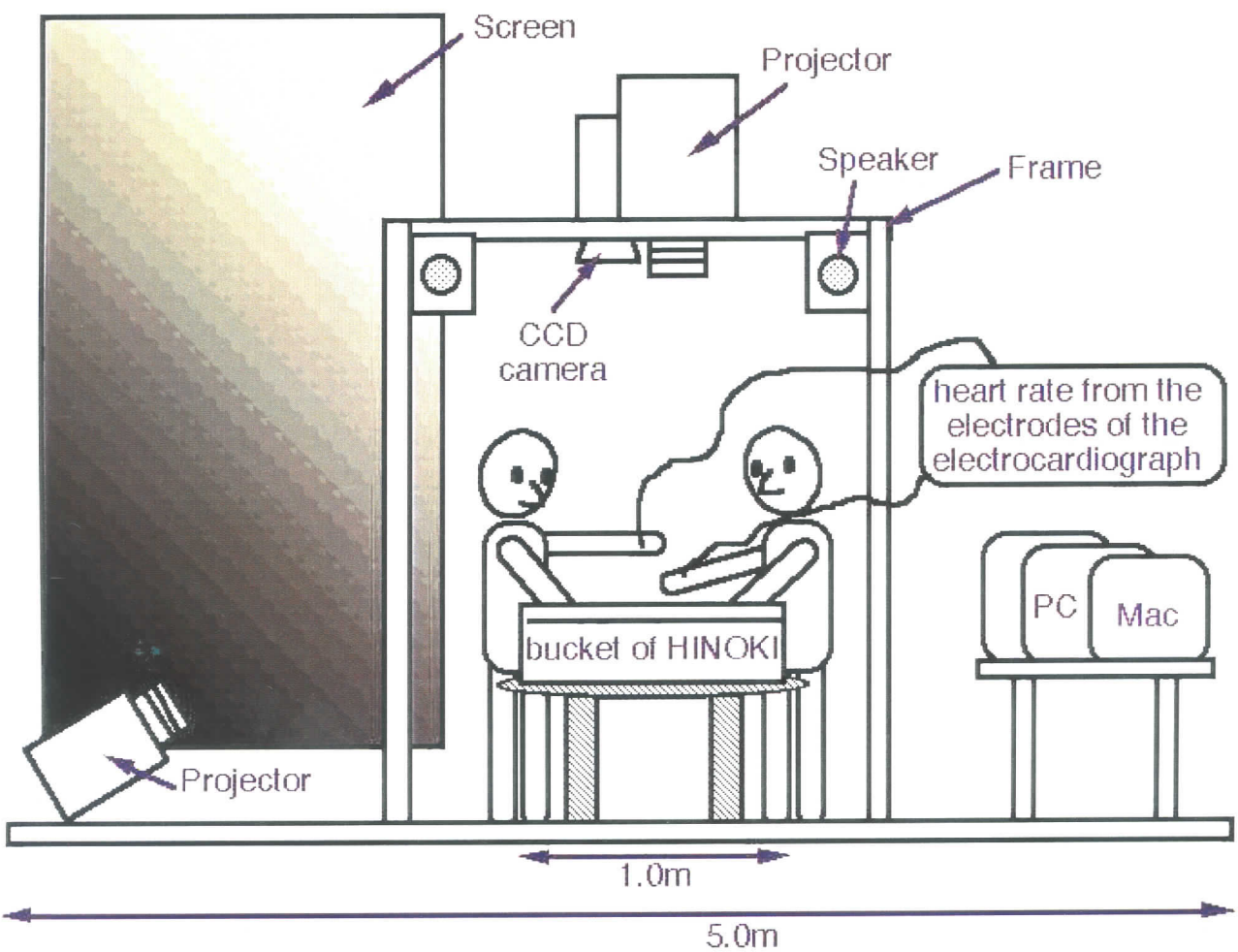
Art Description:

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; 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 while the mind does.



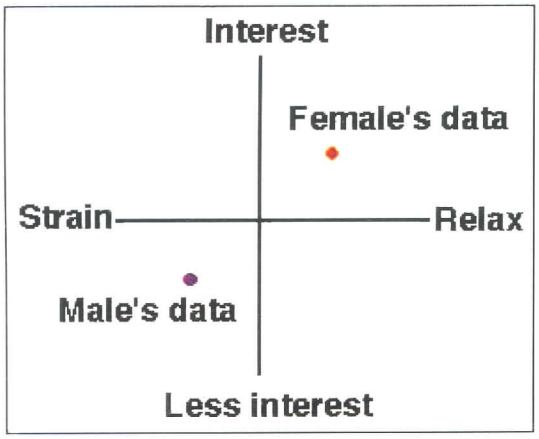
Considering this phenomenon, we propose a means of "touching the heart" in a somewhat Japanese way by measuring the heartbeat of the "honest" body and using other technologies to reveal a new code of non-verbal communication from a hidden dimension in society. We call this "techno-healing art."

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 are mapped on the model. The synchronization interaction model reveals the communication codes in the hidden dimension 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. 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.

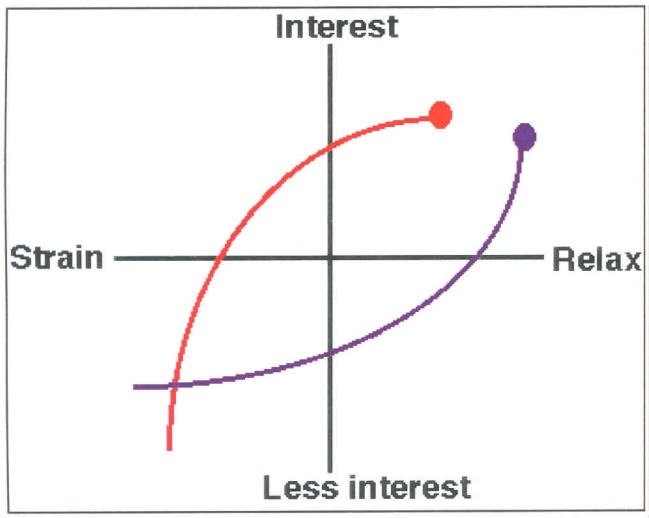


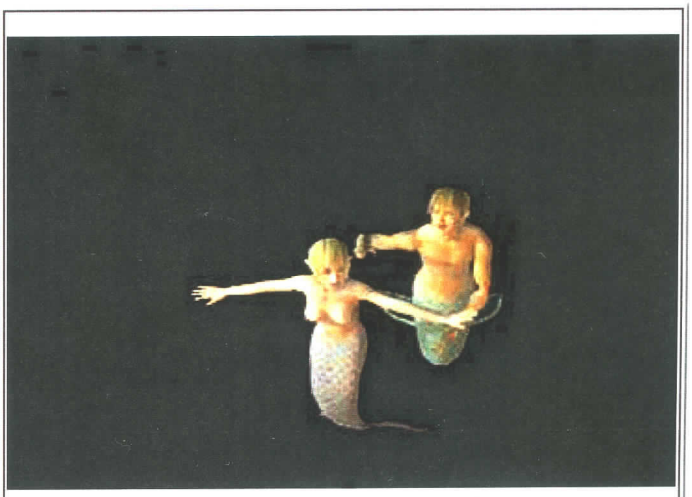
Synchronization interaction model

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 synchronity interaction model reveals the communication codes in the hidden dimension that do not appear in our superficial communication.



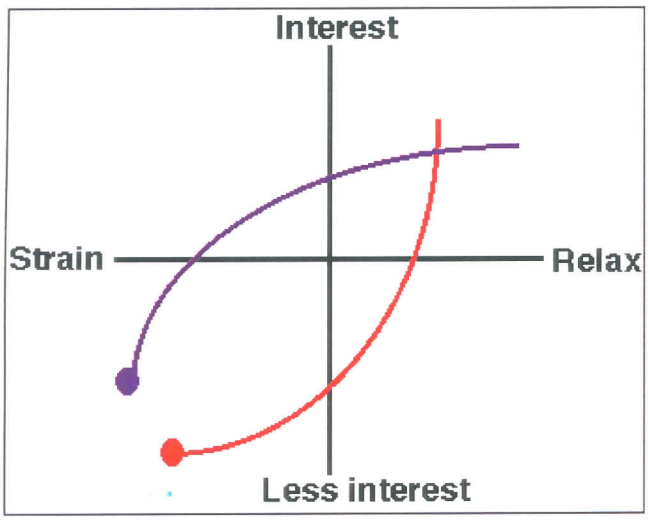
For example, (1) When both people are in the domain where they are highly relaxed and interested, they are considered synchronized. An animation is generated in which, for example, their CG-reactive embodiments join hands in brotherhood or enjoy friendly actions.

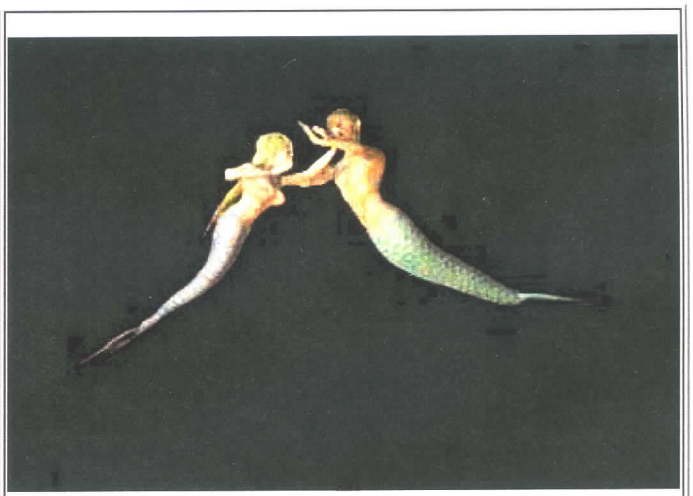




If you click here, QuickTime movie start

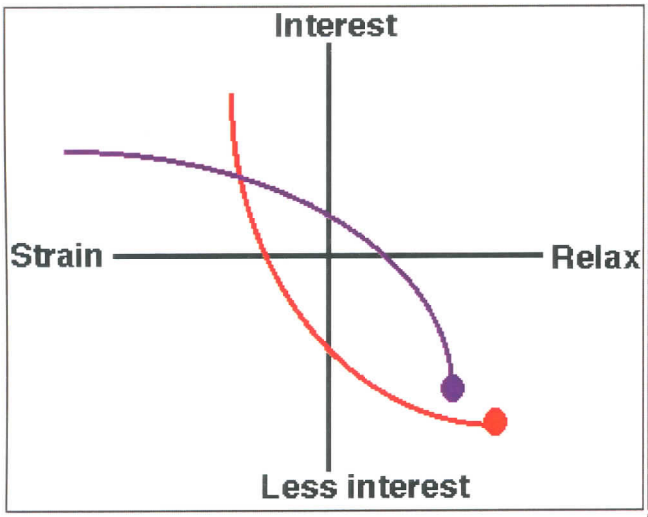
(2) When both people are in a situation where they are highly strained and less interested, unfriendly communication is generated. An animation is generated in which, for example, their CG embodiments quarrel with each other.

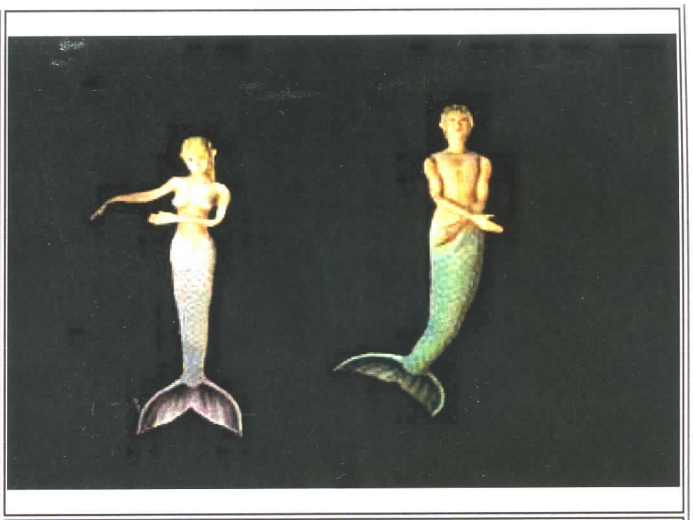




If you click here, QuickTime movie start

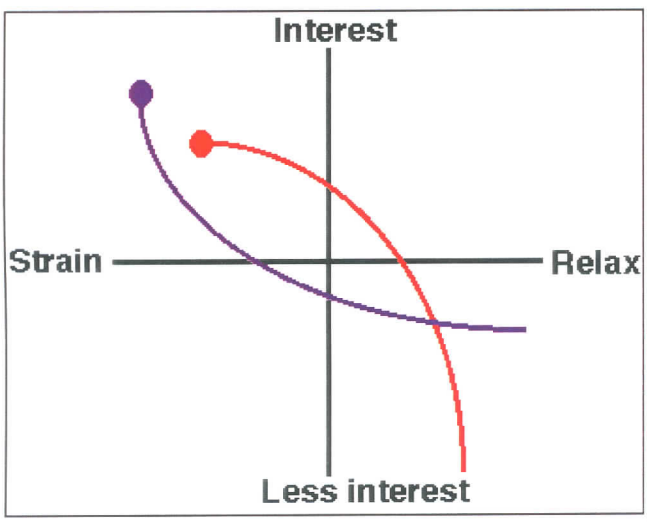
(3) When both people are in the domain where they are highly relaxed and less interested, they are considered, "going their own ways". An animation is generated in which, for example, their CG embodiments do not interfere with each other.

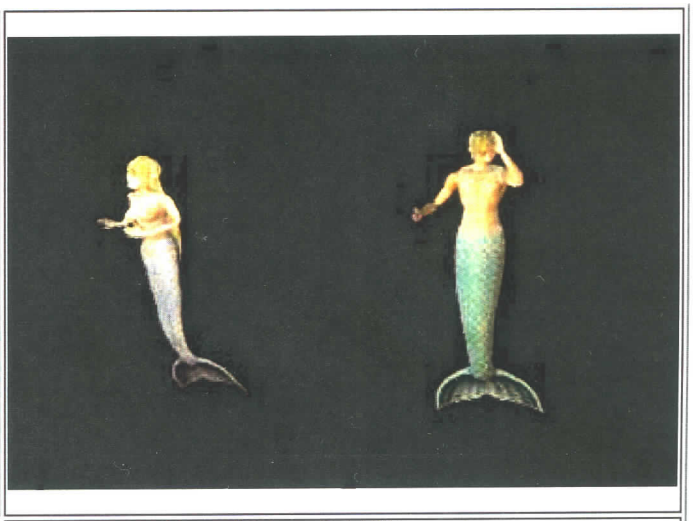




If you click here, QuickTime movie start

(4)when two persons are in a situation where they are highly strained and highly interested, they are assumed to have stress and feelings of shyness. An animation is generated in which, for example, their CG-reactive embodiments behave shyly.





If you click here, QuickTime movie start

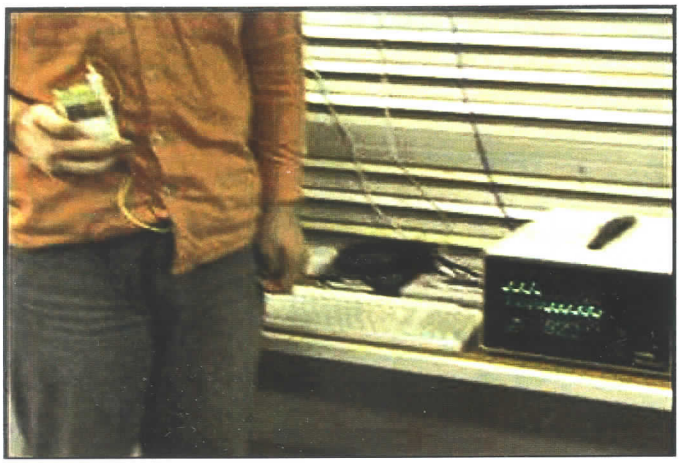
In this way, new codes of non-verbal communication that can't be seen in face-to-face communication are found through the CG of the embodiments.

Hand recognition by using a range sensor

A person's hand is recognized by using a range sensor. The range data of an image is generated from two images inputted from two cameras. A contrast of the image of the hand on the display represents the distance from the cameras; white means that the distance is short and black means it is long.

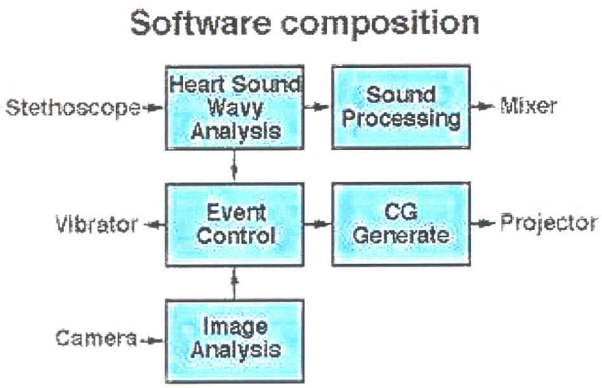
Electrocardiograph

A personÕs heart rate is measured by putting the electrocardiographÕs electrodes on his body. The heart rate is sent to a PC connected to the electrocardiograph via RS232C and is mapped on the synchronity interaction model depending on the heart rate.



Software

Heart Sound Wavy analysis is used to analyze the input data and to send the data to Event Control as event data and to Sound Processing as MIDI commands. Event Control sends several commands to CG Generate if some CG needs to be changed depending on the heart sound data. CG Generate creates CG based on these commands and outputs the CG. Sound Processing processes the sound data as required and then outputs it. Image Analysis analyzes the image data fed from a camera and the relational information of the hand, and the CG displayed is sent to Event Control. Event Control sends some commands to CG Generate if some CG needs to be changed depending on the data. It also sends, when necessary, the changes of the sound data as MIDI commands to Sound Processing or operates vibrators.



Floor plan

For installation, a space with 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 that is filled with water is placed in the center of the installation,. Two persons, fitted with a stethoscope, experience non-verbal communication by touching their CG embodiments in the bucket. The synchronity based on the heart rate from the electrodes of the electrocardiograph is calculated by the PC, and the PC generates an arbitrary feeling in a CG form. The hand movements of the two persons are caught by an installed camera and an image analysis for the data is performed. In accordance with the synchronity interaction model, the CG embodiment either follows the movement of the hand of the partner with high synchronity or goes away from the hand of the partner with low synchronity. When one touches the CG embodiment of the partner, a vibrator gives him a simulated feeling of touch. The stethoscope measures the sound of the heart, which is processed by the PC and outputted.

NAOKO TOSA tosa@mic.atr.co.jp

TRIP

1985
8min.
STEREO, COLOR

Collection

O Art Museum	TOKYO (JAPAN)
Japan Film Culture Center	TOKYO (JAPAN)
Toyama Prefectural Modern Museum	TOYAMA (JAPAN)
Nagoya City Modern Museum	NAGOYA (JAPAN)
American Film Association	N.Y. (U.S.A.)

GIF Animation

PRIZE

1985 Accepted for SCAN '85 Autumn
(Video Gallery SCAN TOKYO)
Accepted for 2nd International Biennal Video CD 85
(Ljubjana YUGOSLAVIA)
Commendable work for NICOGRAPH'85
(Ikebukuro SunShine City Culture Hall TOKYO)

1986 NCGA (National Computer Graphics Association)'86
3rd prize for Independent Artist Section
(Anaheim California U.S.A.)
SIGGRAPH '86 Animation Screening Section
(Dallas Texas U.S.A.)
A Bronze Prize at The International High technology Art Exhibition
(Ikebukuro Sun Shine City Culture Hall TOKYO)
A Special prize at the Japan Image Festival
(TOKYO)

1987 Excellent Prize at the 14th OMNI Art Contest
(Japanese version of OMNI sponsored by bunsha)
Second prize at the American Film & Video Festival; Video Art Section
(New York U.S.A.)
Accepted for The 21st Annual New York Film-Video Exposition
(Metropolitan Art Museum N.Y. U.S.A.)
Accepted for the San Francisco International Film Festival; Golden Gate Awards
(San Francisco U.S.A.)
Accepted for Video Culture Canada '87
(Toronto Canada)

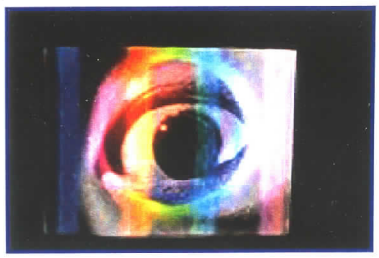
EXHIBITIONS

1986 Invited to and Exhibited at the Monveliarl International Video Festival
(Monveliarl FRANCE)
Invited to and Exhibited at the Japanese Video Art Exhibition; Scanners
(Air Gallery U.K.)
Invited to and Exhibited at the 1st Hot House Exhibition
(Tokorozawa Seibu TOKYO)

1987 Invited to and Exhibited at "Art on Computer"
(O Art Museum TOKYO)

1989 4th Contemporary Art Festival - Image of Today
(Toyama Prefectural Modern Art Museum JAPAN)
TEAM VIDEO GALLERY
(World Design Exhibition Shiratori Hall Nagoya JAPAN)
New Generation Computer Graphics Exhibition
(Kawasaki City Museum JAPAN)

NAOKO TOSA tosa@mic.atr.co.jp



Pleasure

1986
3min.
STEREO. COLOR

GIF Animation

Collection

Takamatsu City Museum TAKAMATSU, KAGAWA (JAPAN)
Nagoya Prefectural Modern Museum NAGOYA (JAPAN)

EXHIBITIONS

1984 Play [MORAL] of Group,NOISE, Image Section
*Superintended, Script and Direction by Koharu Kisaragi
(Studio R TOKYO)

1985 [MORAL 2nd] *Script and Direction by Koharu Kisaragi
(Shibuya Parco Part3 Space Part 3 TOKYO)
[ISLAND] *Script and Direction by Koharu Kisaragi
(Shochiku Benisan pit TOKYO)
[ISLAND;Yokohama] *Script and Direction by Koharu Kisaragi
(Kanagawa Young People Center Hall JAPAN)

1986 Dancing Voice in SEED [Urban Life]
(Shibuya Seibu SEED Hall TOKYO)
[SAMSA] *Script and Direction by Koharu Kisaragi
(FM Tokyo Hall TOKYO)
Invited to and Exhibited at the '86 Osaka International PlayFestival
[MORAL 3rd] *Script and Direction by Koharu Kisaragi
(Kintetsu small theater Osaka JAPAN)

NAOKO TOSA tosa@mic.atr.co.jp

An Expression

1985
9min.
STEREO, COLOR

GIF Animation (1.4MB)

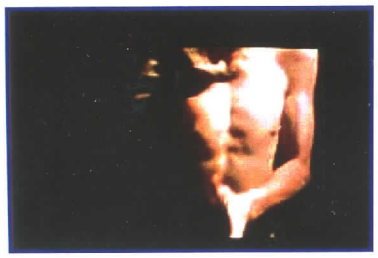
Collection

Nagoya Prefectural Modern Museum NAGOYA (JAPAN)

Exhibition

1986 Invited to and Exhibited at "NEW VIDEO JAPAN"
(Museum of Modern Art U.S.A.)
(Long Beach Museum of Art)
Invited to and Exhibited at the Australian National Broadcasting Systems Program;
International Video Art Section
(AUSTRALIA)
Video Cocktail 3
(Hara Art Museum TOKYO)
Invited to and Exhibited at the 1st Hot House Exhibition
(Tokorozawa Seibu TOKYO)
1988 Invited to and Exhibited at the EXPERIMENTAL MEDIA FESTIVAL; MIAMI WAVES
(Miami-Dade Community College U.S.A.)
1989 4th Contemporary Art Festival - Image of Today
(Toyama Prefectural Modern Art Museum JAPAN)
TEAM VIDEO GALLERY
(World Design Exhibition Shiratori Hall Nagoya JAPAN)
New Generation Computer Graphics Exhibition
(Kawasaki City Museum JAPAN)

NAOKO TOSA tosa@mic.atr.co.jp



ECSTACY

1986
7min.
STEREO, COLOR

GIF Animation (1.3MB)

Collection

National Art Museum	OSAKA (JAPAN)
Toyama Prefectural Modern Museum	TOYAMA (JAPAN)
Nagoya City Modern Museum	NAGOYA (JAPAN)

PRIZE

1987 The best producer prize for International Image & Music Prize sponsored by Fuji TV
(National Yoyogi ground First gymnasium TOKYO)
Excellent Prize at the 14th OMNI Art Contest
(Japanese version of OMNI sponsored by bunsha)
Accepted for Art Document '87
(Tochigi art museum JAPAN)
Accepted for the Film & Video Festival; Video section
(New York U.S.A.)
Accepted for SIGGRAPH '87
(Anaheim California U.S.A.)
Accepted for Video Culture Canada '87
(Toronto Canada)
1988 Accepted for the BACA's 21st Annual Film-Video Festival
(Jefferson Market Library N.Y. U.S.A.)

EXHIBITIONS

1986 Invited to and Exhibited at the Japanese Video Art Exhibition; Scanners
(Air Gallery U.K.)
Invited to and Exhibited at the Camerino International Video Festival
(Camerino ITALY)
Invited to and Exhibited at the 1st Hot House Exhibition
(Tokorozawa Seibu TOKYO)
1987 Invited to and Exhibited at the Festival International de Film Et Video de Montreal
(CANADA)
Invited to and Exhibited at "Art on Computer"
(O Art Museum TOKYO)
Invited to and Exhibited at the International High Technology Art Exhibition
(NS bld. 1F TOKYO)
ACM/SIGGRAPH TRAVELING ART SHOW; 25 years of Computers in the Arts
(California State University L.A. U.S.A.)
Invited to and Exhibited at "Image de Future '87"
(Montreal CANADA)
Invited to and Exhibited at the Sapporo High Tech Art Exhibition
(Art Forest in Sapporo, Art Lobby Hokkaido JAPAN)
Video Select '87
(Hokkaido Prefectural Modern Art Museum JAPAN)
1988 Invited to and Exhibited at the Brisbane International Leisure Center "Japan Techno Plaza"
(Brisbane AUSTRALIA)
Video Festival; Japan Now - Sweden Now ppppp
(Kalturhuset SWEDEN)
Invited to and Exhibited at the International High Technology Art Exhibition
(Tsukashin Hall Kobe JAPAN)
Invited to and Exhibited at Fukui International Video Biennal
(Phoenix Plaza Fukui JAPAN)
EXPERIMENTS IN ANIMATION Exhibition
(O Art Museum TOKYO)
4th Contemporary Art Festival - Image of Today
(Toyama Prefectural Modern Art Museum JAPAN)
TEAM VIDEO GALLERY
(World Design Exhibition Shiratori Hall Nagoya JAPAN)
New Generation Computer Graphics Exhibition
(Kawasaki City Museum JAPAN)

NAOKO TOSA tosa@mic.atr.co.jp

GUSH!

1989
7min.
STEREO, COLOR

GIF Animation

Collection

Nagoya Prefectural Modern Museum NAGOYA (JAPAN)

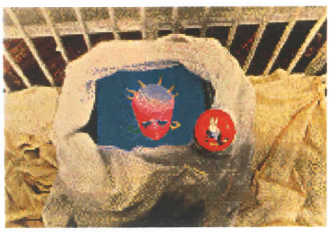
PRIZE

1989 Accepted for ARTEC '89 : First Nagoya International Biennal;
Domestic Public contribution Invitation
(Nagoya City Science Museum JAPAN)
1990 Accepted for SCAN '90 Video Art; New works of Public contribution
(Heineken Village Gallery TOKYO JAPAN)

EXHIBITIONS

1989 4th Contemporary Art Festival - Image of Today
(Toyama Prefectural Modern Art Museum JAPAN)
TEAM VIDEO GALLERY
(World Design Exhibition Shiratori Hall Nagoya JAPAN)
New Generation Computer Graphics Exhibition
(Kawasaki City Museum JAPAN)
1990 Invited to and Exhibited at the Monveliarl International Video Festival
(Monveliarl FRANCE)
Invited to and Exhibited at the Locarno International Video Festival
(Locarno SWITZERLAND)

NAOKO TOSA tosa@mic.atr.co.jp



Neuro Baby

1992-1994

Showing

<1993>

Ars Electronica 1993 "Artificial Life"
Linz Austria

SIGGRAPH "Machine Culuture"
Anaheim U.S.A.

A-Life world /Tokyo International Museum
Tokyo Japan

<1994>

Interactive Art Exhibition /SUM Museum (Tenpozan Museum)
Osaka Japan

Technical Detail

Paper

Linked Page

http://userwww.sfsu.edu/~swilson/emerging/artre482.AI.html
http://www.dlf.de/computer/ck980620.html
http://www.nmao.go.jp/tj/tenranj_31.html
http://www.dnp.co.jp/museum/nmp/nmp_b/information/Sep10_j.html
http://www.sankei.co.jp/databox/paper/9607/paper/0709/miraisi.html
http://www.aec.at/fest/fest93/tosa.html
http://geo.de/wissen/96/01/FreundCyberspace.html

HOME
NAOKO TOSA tosa@mic.atr.co.jp

Neuro Baby

Naoko Tosa
ATR Media Integration & Communications Research Laboratories
Seika-cho Soraku-gun Kyoto, Japan
Phone: +81 774 95 1427
tosa@mic.atr.co.jp
http://www.mic.atr.co.jp/~tosa/

An automatic facial expression synthesizer that responds to expressions of feeling in the human voice.

I created a new creature or a piece of work that can live and meaningfully communicate with modern, urban people like ourselves, people who are overwhelmed, if not tortured, by the relentless flow of information, and whose peace of mind can only be found in momentary human pleasures. Neuro Baby was born to offer such pleasures.

The name "Neuro Baby" implies the "birth" of a virtual creature, made possible by the recent development of neurally based computer architectures. Neuro Baby "lives" within a computer and communicates with others through its responses to inflections in human voice patterns. Neuro Baby is reborn every time the computer is switched on, and it departs when the computer is turned off. Neuro Baby's logic pattern are modeled after those of human beings, which make it possible to simulate a wide range of personality traits and reactions to life experiences.

Neuro Baby can be a toy, or a lovely pet ---or it may develop greater intelligence and stimulate one to challenge traditional meanings of the phrase "intelligent life". In ancient times, people expressed their dreams of the future in the media at hand, such as in novels, films, and drawings. Neuro Baby is a use of contemporary media to express today's dreams of a future being.

This work is the simulation of baby, born into the "mind" of the computer. Neuro Baby is a totally new type of interactive performance system, which responds to human voice input with a computer-generated baby's face and sound effects. If the speaker's tone is gentle and soothing, the baby in the monitor smiles and responds with a sad or angry expression and voice. If you try to chastise it, with a loud cough or disapproving sound, it becomes needy and starts crying. The baby also sometimes responds to special events with a yawn, a hiccup, or a cry. If the baby is ignored, it passes time by whistling, and responds with a cheerful "Hi" once spoken to.

The baby's responses appear very realistic, and may become quite endearing once the speaker becomes skilled at evoking the baby's emotions. It is a truly lovable and playful imp and entertainer. In many ways, it is intended to remind speakers of the life-like manner of the famous video-computer character Max Headroom.

Two major technologies were combined to create this system: voice analysis and the synthesis of facial expressions.

Voice analysis was performed by a neural network emulator that converted the voice input wave patterns into "emotional patterns" represented by two floating point values. The neural network has been "taught" the relationship between inflections in human voices and emotions patterns contained within those inflections. During interaction with the baby the emotional patterns found in the observer's speech are continuously generated.

During the translation stage, the two values for emotional patterns are interpreted as an X-Y location on an emotional plane, onto which several types of emotional patterns are mapped. For example, "anger" may be located on the lower left of such a plane, while "pleasure" would be located on the upper right of the same plane. Each emotional pattern corresponds to a paired facial expression and a few seconds of voice output.

During the performance, the facial expression is determined be interpolating the shape, position, and angle of facial parts, such as eyes, eyebrows, and lips. These parts were pre-designed for each emotional reaction. One FM TOWNS, Fujitsu's multimedia personal computer, is used for voice analysis, another FM TOWNS is used for voice generation, and a Silicon Graphics IRIS 4D is used for image synthesis.

Network Neuro Baby

1995

Showing

<1995>

Insutitute of Industrial Sciences, University of Tokyo
Roppongi Tokyo Japan

ACM SIGGRAPH '95
Los Angels USA

Technical Detail

Paper

Linked Page

http://kslab-www.pi.titech.ac.jp/Juten/hhashimoto/hashimoto.html

HOME
NAOKO TOSA tosa@mic.atr.co.jp

Network Neuro-Baby with robotics hand

(An automatic facial expression systhesizer that responds to espressions of feeling in the human boice and handshake)

Naoko Tosa* Hideki Hashimoto Kaoru Sezaki** Yasuharu Kunii** Toyotoshi Ymaguchi** Kotoro Sabe** Ryosuke Nishino** Hiroshi Harashima*** Fumio Harashima****

*ATR Media Integration & Communications Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

**Insutitute of Industrial Sciences, University of Tokyo
7-22-1 Roppongi, Minato-ku, Tokyo 106 Japan

***Department of Electrical Enginnering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113 Japan

Abstract

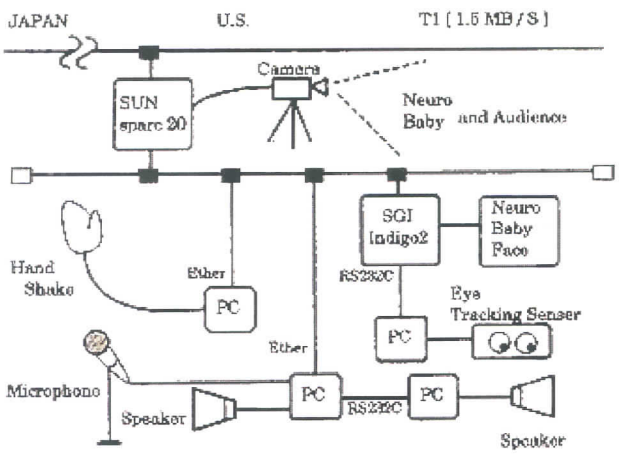
Neuro-Baby is a totally new type of interactive performance system which responds to the human voice with a computer-generated baby face and sound effect. Emotin space model is employed to categolize the feeling of the speaker. To recognize the human voice we used a nueural network which has been taught the relationship between a set of digitized wave patterns and the location of several emotion types in the emotion space. The facial expression is systhesized continuously according to the location which the neural network generates. The flexible design of NB is pollible by changing the facial design, the layout in the emotion space, sensitiviity to the transition of the feelings or the teaching pattern for the neural network.

By networking NB's, we can enjoy a non-verbal communication with each other. Such a Networked NB's will help the mutual understanding, absorption of cultural gap asa well as international cultural exchange very much. The first result will be demonstrated in 1995, by connecting two NB's between Japan and USA. The networking issues concerning such a systen is also addressed.

Introduction

A new creture has benn born!! This creature can libe and meaningfully communicate with modern, urban people lide ourselves, people who are overwhelmed, if not tortured be the relntless folw of information, and whose peace of mind can only be found momentary human pleasures. NB was born to offer such pleasures. The name "NB" implies the "birth" of a virtual creature, made possible by the recent development of neurally based computer architectures. NB "lives" within a computer and communicates with others through its responses to inflections in human voice pattens. NB is reborn every time the computer is switch on, and it departs when the computer is turned off. NB's logic pattens are modeled after those of humans beings, which make it possible to simulate a wide range of personality traits and reactions to life experiences. NB can be a toy, or a lovely pet - or it may develop greater intelligence and stimulate one to challenge traditional meanings of the phrase "intelligent life." In ancient times, people expressed their dreams of the future in the media at hand, such as in novels, films, and drawings. NB is a use of contemporary media to express today's dream of a future being. [1]

System configuration for Network Neuro-Baby with robotics hand



Basic characteristic of Neuro-Baby and its interaction with the external world

If the speaker's tone is gentle and soothing, the baby in the monitor smiles and responds with a pre-recorded laughing voice. If the speaker's voice is low or threatening, the baby responds with a sad or angry expression and voice. If you try to chastise it, with a loud cough or disapproving sound it becomes sad and starts crying. The baby also sometimes responds to special events with a yawn, a hiccup, or a cry. If the baby is ignored, it passes time by whistling, and responds with a cheerful "Hi" once spoken to. The baby's responses appear very realistic, and may become quite endearing once the speaker becomes skilled at evoking the baby's emotions.

Figure 2. Figure3 shows the general model of the NB from human input via recognition mapping R to a state in the emotion model, and then via the expression mapping E to the output.

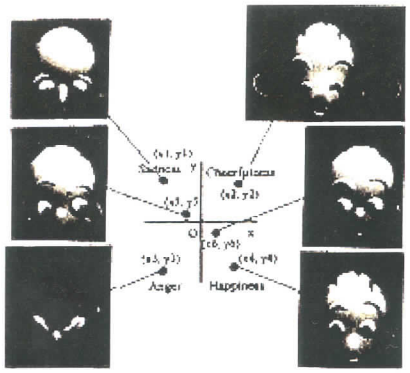


Fig. 2. An assignment of the model faces

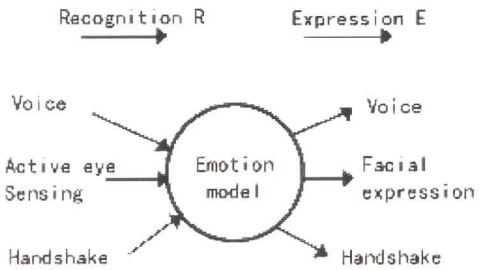


Fig. 3. Processing model of the NB

Customization and Learning of the relationship between voice and emotion

The principle function of NB is to make a map describing emotional responses evoked by voice input so that speakers can feel these emotional responses naturally and comfortably. The emotional responses are expressed by using x-y axis coordinate shown in Figure 4. We call the coordinate as an emotional model. A point (x,y) corresponds to an action which NB performs to express his/her response. The coordnate of emotional model has been changed by a neural network with a set sampled data such as sadness, cheerfulness, anger and happiness.[2] NB has several types of emotional models and speakers can select one among them in accordance with speaker's characteristics. This selection is dependent on first input voice and reaction of handshaking machine. It is a kind of customization and realizes more delicate responses.

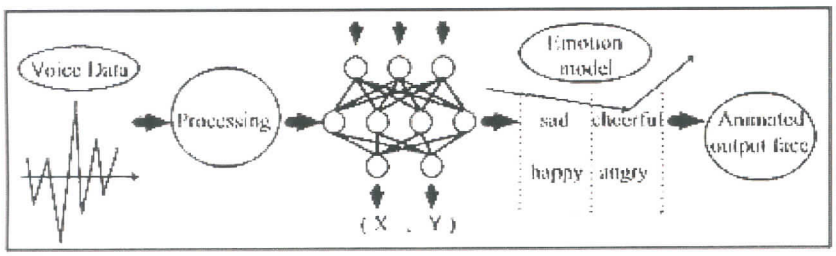


Fig. 4.

Handshaking Device on Network Neuro-Baby

Handshaking Device (HSD) is an interface device which speakers can communicate with NB physically. The HSDs with NB are placed in Japan and USA so that people can communicate each other physically with HSDs through NB. The structure of HSD is shown in Figure 5. It can make force sensation to an operator and measure force pressure by pressure sensor. The HSD is assumed as a right hand and then the operator can feel existence of human through the force sensation generated by HSD. The other HSD is also grasped by other people. Those HSDs can send and receive force sensation through information network. The HSD is used as an input device to NB instead of Key-Board when the emotional model is customized. Figure 6 shows system structure of HSD. The HSD is composed by 2 linear motors (AM20), and position sensor and force sensor are implemented to measure force from an operator. Those information are connected to a host computer via transputer mother board and i860 through 20Mbps link The i860 is used to real time control. The host computer is connected to a host workstation (SSIO) to communicate with the other HSD through information network (ATM and optical fiber).

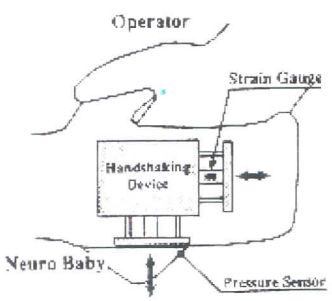


Fig. 5. Handshaking Device

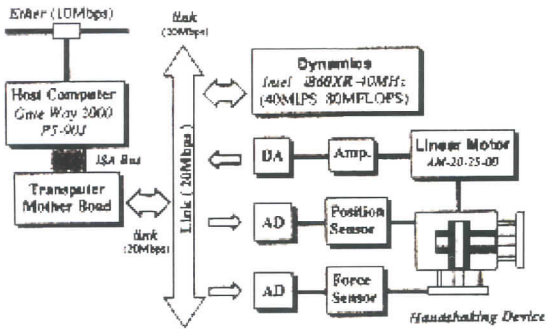


Fig. 6. System of Handshaking Device

Active Eye Sensing System

Active Eye Sensing System for NB is used to get information about speakers' face position so that NB can look at speakers. It means finding the most similar face among ones in the camera image by using template matching. In future it can recognize facial expression by understanding images. Figure 7 shows the active eye sensing system. Each camera has two servo motors and then get two-degree-of-freedom such as yaw and pitch. The stereo camera system can identify the pose of moving object. The image is transformed into digital by Video Module and transferred to Tracking Module through VMEbus. In the tracking module three frames are stored and the motion between frames is estimated. These modules are controlled by VME master transputer which also calculate the pose of moving object. The pose information is used to control servo motors and transferred to NB through the host PC.

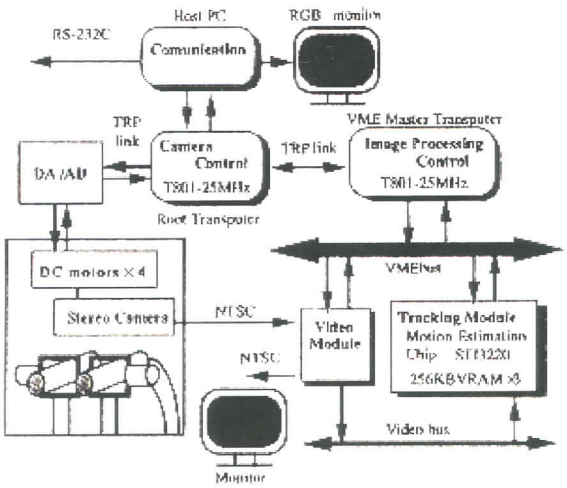


Fig. 7. Active Eye Sensing System for Neuro Baby

Neuro-Baby as Network Sensitive Agent

By networking, various new issues will happen. Since a network is subject to error and delay, the compensation of these effect is to be solved. Fore "conventional" media as image and voice, many techniques for it appeared in literature [3]. However, since networked NB is totally a new application, there exist no technique for delay or error compensation. Therefore, we developed new inter media and inter media synchronization technique suitable for the handshaking. This technique may also be used in general teleoperation systems, also a scaling technique is considered because the network might be one with long delay and severe packet loss.

Application:International Cultural Exchange

A networked NB can be used to help improving international cultural exchange and absorbing the cultural gap by customizing the NB at each site. Figure 8 shows two communication partners, one in Japan and one in the U.S., communicating via two NB's. The NB in Japan is customized for a Japanese user with appropriate recognition and expression mapping, whereas the NB in the U.S. is customized for its user with different typical mappings. The NBs communicate their emotional state over the network, which is then expressed to the individual user in both countries in an understandable form with customized expression mappings. A communication setup like this can help reduce cultural differences between differing ways of communication and expression of feelings.

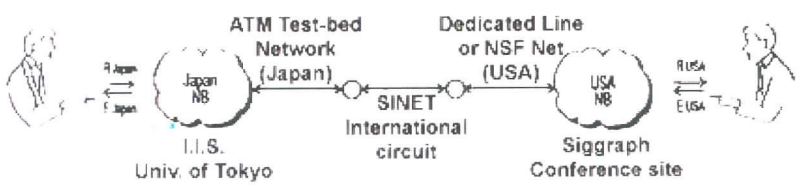


Fig. 8. Network based Neuro-Baby

Demonstration

As shown in Figure 8, we use ATM test-bed network from I.I.S. to the gateway of SINET international circuit which is located in Chiba, Japan. Then, SINET reaches to Stockton, CA. For the lind between Stockton and Confernce site, we will use either dedicated line or NSFnet. The demonstration for SIGGRAPH '95 includes all above characteristics of the improved NB. Especially the demondtration of ineternational cultural exchange will be exhibited. Communication will take place between the SIGGRAPH site (Los Angels) and I.I.S. Univ. of Tokyo. Further it is to be expected that during the yearly I.I.S. Univ. of Tokyo open house event a large number of

people will try to communicate with the NB.

Acknowledgment

The authors woule like to thank NACSIS staffs and Prof. S. Asano for arranging SINET international circuit. This work was partly supported by Grant-in-Aid for Creative Basic Research (Development of High-Performance Communcation Network for Scientific Researchers). The first version of NB was developed in collaboration with N. Tosa and Fujitsu Lab. Special thanks for ATR Media Integration & Communications Research Laboratories.

Reference

[1] Tosa, N., Murakami, K., kakimoto, M., Sato, S., "Neuro-Character" AAAI '94 Workshop:AI and A-Life and Entertainment

[2] Philips, Sound Ideas, "Sound Effect Library LFF7910 Series," CD1001-CD1028.

[3] See for example, Proc. 6th international workshop on Packet Video, Sep. 1994.

Showing List

HOME

NAOKO TOSA tosa@mic.atr.co.jp

Interactive Poem

1996-1998

Award

1997 L'Oreal prize / L'Oreal Arts and Science Foundation

Showing

<1997>
ISEA97
Chicago
September

DIGITAL MEDIA WORLD (NICOGRAPH/Nikkei)
Makuhari-Messe Japan
December

<1998>
Exhibition: A-Life IV
L.A. U.S.A.
June

Virtual Enveronment '98
Stuttgart, Germmany
June

ACM Multi Media 98 ART demo
Bristol, U.K.
September

Images

GIF Slide Show

Movie

QuickTime Movie

Technical Detail

Paper

Setup requirement

Setup requirement

Floor plan

Floor plan

NAOKO TOSA tosa@mic.atr.co.jp

Interactive Poem System

Naoko Tosa
ATR Media Integration & Communications Research Laboratories
Seika-cho Soraku-gun Kyoto, Japan
Phone: +81 774 95 1427
tosa@mic.atr.co.jp
http://www.mic.atr.co.jp/~tosa/

Ryohei Nakatsu
ATR Media Integration & Communications Research Laboratories
Seika-cho Soraku-gun Kyoto, Japan
Phone: +81 774 95 1400
nakatsu@mic.atr.co.jp
http://www.mic.atr.co.jp/~nakatsu/

Abstract

We propose a new type of speech-based interaction system called "Interactive Poem". Conventional speech-based interaction systems have only focused on the transmission of logical meaning involved in speech. The application of such systems has been restricted to business services such as making reservations or data retrieval. In the Interactive Poem system, however, a human and a computer agent create a poetic world by exchanging poetic phrases, thus realizing Kansei-based communications between computers and humans. This paper first proposes the concept of "Interactive Poem". It then describes the details of the system we have developed, including the software and the hardware configurations as well as the interaction mechanism.

Keywords

Interactive Art, Emotion Recognition, Art & Technology research, New type Mulitimedia contents, poetic interface

Table of Contents

- Introduction
- CONCEPT OF INTERACTIVE POEM
- SOFTWARE CONFIGURATION
- INTERACTIONS
- HARDWARE CONFIGURATION
- CONCLUSION
- Bibliography

Introduction

In human oral communications, sensitivity information, such as emotions and sensitivities, plays a very important role. Sensitive information is sometimes more important than the logical information included in speech. This can be confirmed by the fact that babies start to recognize emotional information before they can recognize logical information in their mothers' voice. In the case of adults, we too can recognize what other people want to say at a deeper level by integrating both logical and sensitivity information included in speech. This is the key to smooth communications. Unfortunately, in the field of AI so far, the focus has been on recognition of only meaning information, while Sensitive information has been neglected as noise.

Based on the above considerations, we started to study how to realize emotion-based communications between computer agents and humans. As a first step toward this aim, we have developed several computer agents such as "Neuro Baby"[4] and "MIC and MUSE"[5]. These are computer characters that are capable of recognizing several emotions included in speech and reacting to them by changing their facial expressions and body motions. Fortunately, these agents have been very successful and have been demonstrated at various exhibitions.

emotion-based communications, however, is only part of emotion-based communications, and we are interested in yet a deeper level of communications. As a next step toward the realization of emotion-based communications between computer agents and humans, we have selected "poem" as a means of communications. There are several reasons for this approach. The main reason is that in a poem not only the meaning of words or phrases but also the rhythms and moods created by their sequence plays an essential role. Therefore, the poem is intended to transmit Sensitivity information such as mood and sensitivity rather than logical information. The second reason is that poems were originally expressed by oral reading rather than in writing. This means that a poem is suitable for interaction between computers and humans. Recently, researchets have shown increased interest in the realization of emotion-based interactions and communications between computers and humans[1][2][3]. However, only few have treated voice communications, despite the fact that voice is an essential means of sensitivity-communications. This is the third reason why we are interested in treating communications based on an uttered poem. This paper first introduces the concept of interactive poem. Then the basic principles of the Interactive Poem system we have developed based on the concept are described. The software configuration and hardware configuration will be described in detail. Finally, a typical installation of the Interactive Poem system will be introduced.

<-- Table of Contents

CONCEPT OF INTERACTIVE POEM

"Interactive poem" is a new type of poem that is created by a participant and a computer agent collaborating in a poetic world full of inspiration, emotion and sensitivity.

In a conventional poem, a poet tries to convey emotions and sensitivity through a sequence of carefully selected words and short phrases or sentences. Because of the magic power unleashed by the words and phrases, people can easily understand the message that a poet wants to express and can thus enter this world created by the poet. However, the world of each poem is static and, therefore, limited by the intentions of the poet because the phrases, sentences and especially their sequence are fixed.

The concept of this interactive poem is based on conventional poetry, but goes beyond traditional limits by introducing the capability of interaction. A participant and a computer agent create a dialogue by exchanging short poetic phrases, and through this exchange produce a new poetic world that integrates the poetic world of the agent with his/her own.

A computer agent called "MUSE", who has been carefully designed with a face suitable for expressing the emotions of a poetic world, appears on the screen. She will utter a short poetic phrase to the participant. Hearing it allows him/her to enter the world of the poem and, at the same time, feel an impulse to respond by uttering one of the optional phrases or by creating his/her own poetic phrase. Exchanging poetic phrases through this interactive process allows the participant and MUSE to become collaborative poets who generate a new poem and a new poetic world.

<-- Table of Contents

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. 1).

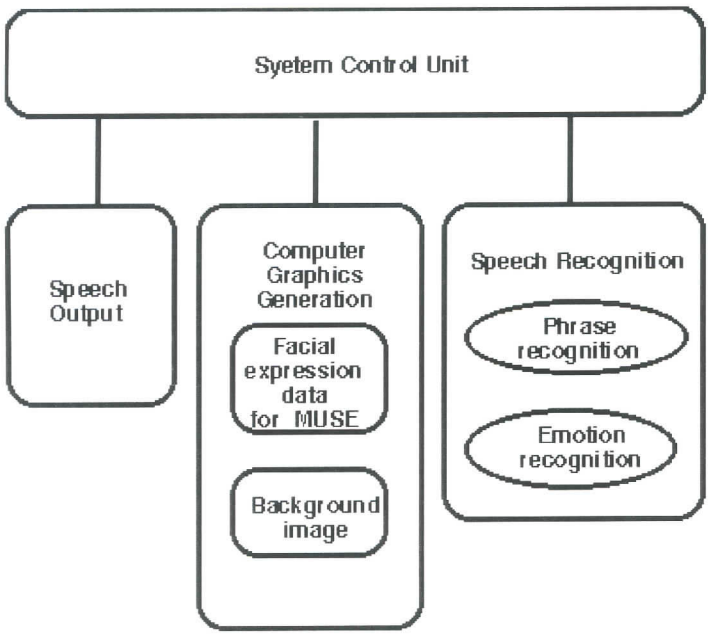


Fig.1 Block diagram of the Interactive Poem System

The system control unit manages behavior of the whole system by utilizing the interactive 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. 2-A).

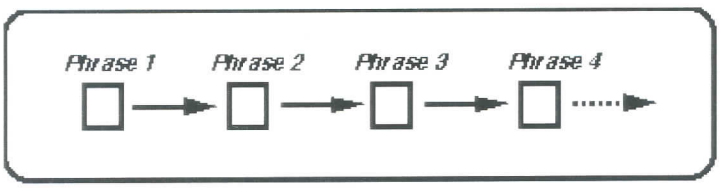


Fig. 2-A Conventional Poem

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. 2-B).

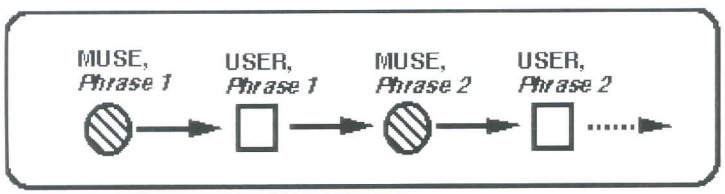


fig. 2-B Construction of the Interactive Poem (a)

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. 2-C). These phrases are carefully created and chosen by taking into account how 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.

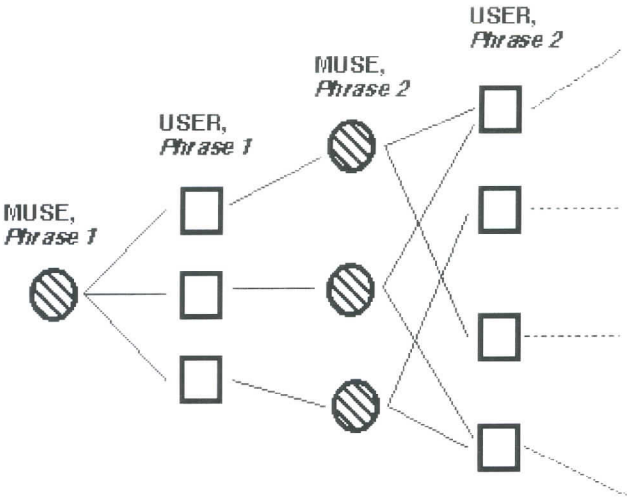


fig. 2-C Construction of the Interactive Poem (b)

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. Each phrase to be uttered is represented in the form of a phoneme sequence and is stored in the lexicon (Fig. 3). 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. 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 (Fig. 4).

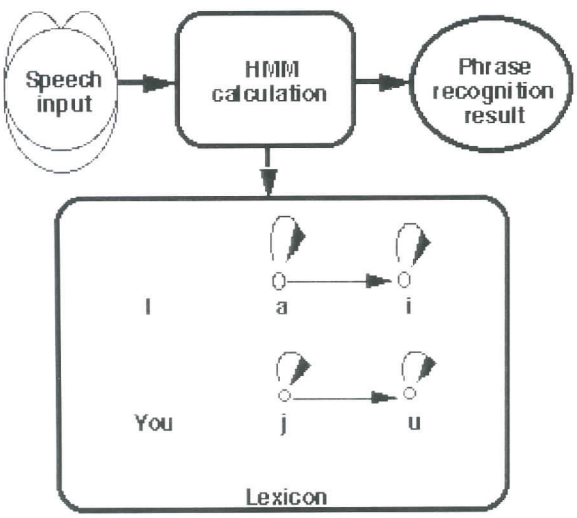


Fig. 3 Phrase recognition

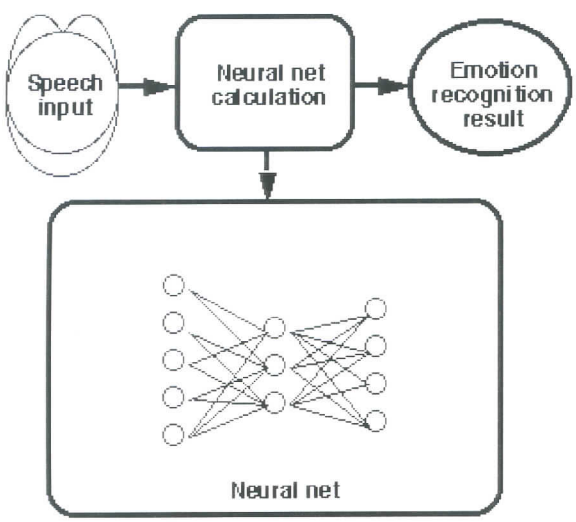


Fig. 4 Emotion recognition

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 change depending on the emotion recognition result and express her reactions to the emotional state of a participant.

These images are represented by keyframe 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. 6).

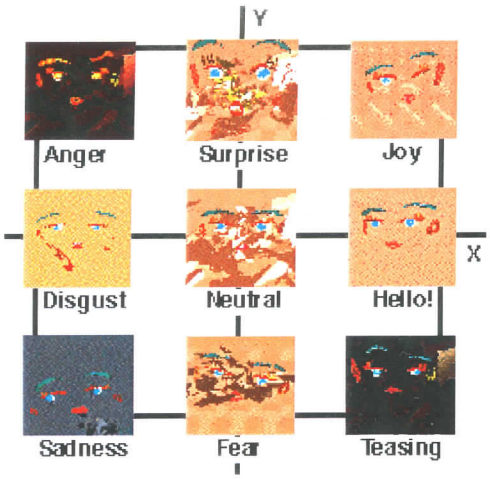


Fig. 5 Muse's emotional expression

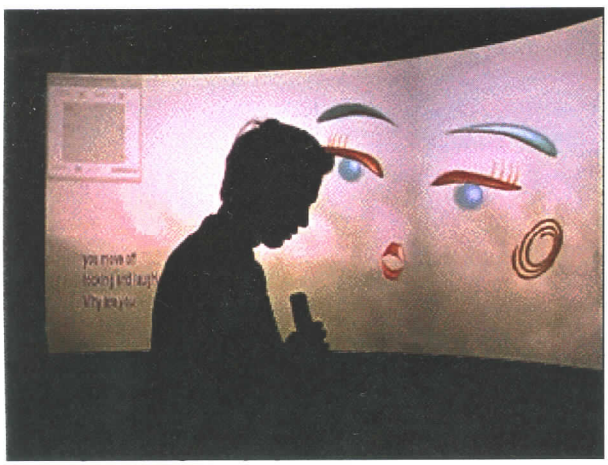


Fig. 6 Interaction between Muse and the participant
Download QuickTime Movie (20sec. 956KB)

<-- Table of Contents

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.

<-- Table of Contents

HARDWARE CONFIGURATION

This system 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. Regardless, the emotion recognition function can produce a result. In addition, 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. 7).

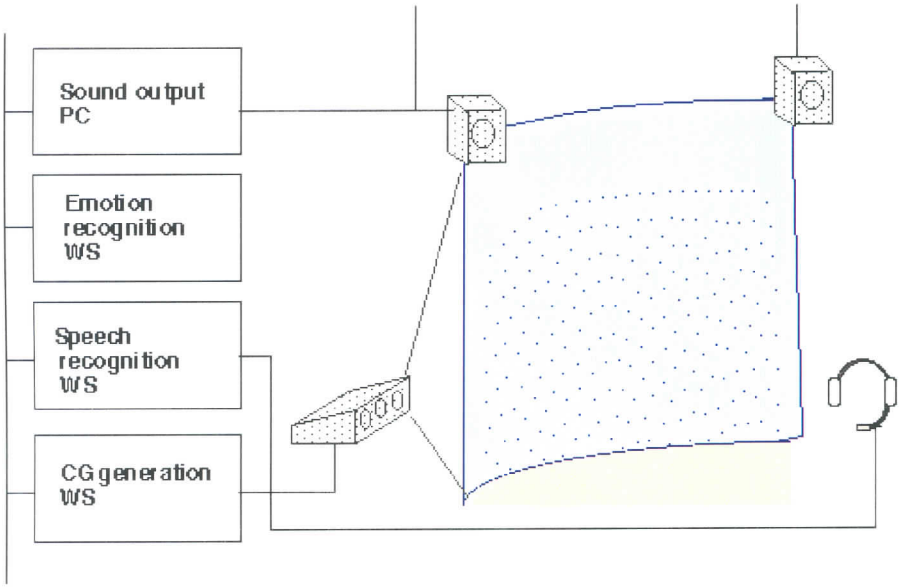


Fig.7 Interactive Poem hardware configration

<-- Table of Contents

CONCLUSION

In this paper we have proposed a new concept based on Sensitivity interaction called "Interactive Poem." "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. This system was realized by the collaboration between an artist and an engineer. From the artistic point of view, the production of a computer poet called "MUSE," especially its facial expressions, is a key issue. The creation of background images that fit the mood of the poem is another key issue. From the technology point of view, speech recognition and emotion recognition play key functions. By integrating these two different types of skills and talents, we could produce a new system that can be considered both a new interactive art medium and a new emotion-based interaction system.

<-- Table of Contents

Bibliography

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

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

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

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

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

<-- Table of Contents

HOME

NAOKO TOSA tosa@mic.atr.co.jp

Interactive Poem

Setup requirement

Hardware

SGI Indigi2 Maxam Impact(R10000)	1
SUN Sparc20	1
PC	2
Laser Disk Recoder	1
RGB Projecter and 200 inch screen	1
Speaker	1
Mixer	1
Microphone	1

Software

Orignal software

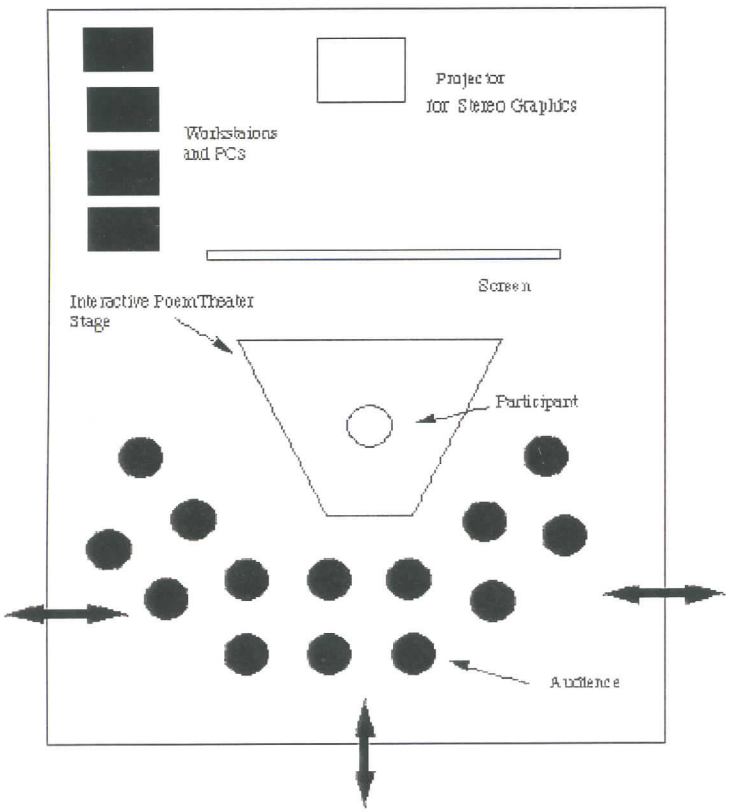
HOME

NAOKO TOSA tosa@mic.atr.co.jp

Interactive Poem

Floor Plan

This figure shows the floor plan of the interactive poem system. A participant stands in front of a large screen and interacts with MUSE by exchanging poetic phrases with her. Several audience members may watch and enjoy the ongoing poem production process.



Floor plan of the Interactive Poem System

HOME

NAOKO TOSA tosa@mic.atr.co.jp

Interactive Theater
"Romeo & Juliet in Hades"

1998-1999

Concept

Slide Presentation

Award

<1998>
Japan Telecom Prize / The Telecommunications Advancement Foundation

Showing

<1999>
International Berlin Film Festival Newmedia Division "Transmediale '99"
Podewil Berlin Germany
Feb.

Technical Detail

Paper

Images

Interactions
Selected Scenes

Movie

QuickTime Movie (22sec. 1MB)

Linked Page

http://www.kyoto-np.co.jp/kp/topics/99jan/25/04.html
http://www.kyoto-np.co.jp/kp/people/gakken/gakken07.html
http://www.crcpress.com/catalog/1825.htm

NAOKO TOSA tosa@mic.atr.co.jp

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

Ryohei Nakatsu
ATR Media Integration & Communications Research Laboratories
Seika-cho Soraku-gun Kyoto, Japan
Phone: +81 774 95 1400
nakatsu@mic.atr.co.jp
http://www.mic.atr.co.jp/~nakatsu/

Naoko Tosa
ATR Media Integration & Communications Research Laboratories
Seika-cho Soraku-gun Kyoto, Japan
Phone: +81 774 95 1427
tosa@mic.atr.co.jp
http://www.mic.atr.co.jp/~tosa/

Takeshi Ochi
ATR Media Integration & Communications Research Laboratories
Seika-cho Soraku-gun Kyoto, Japan
Phone: +81 774 95 1494
ochi@mic.atr.co.jp
http://www.mic.atr.co.jp/~ochi/

Abstract

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

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

Keywords

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

Table of Contents

- Introduction
- Outline of Interactive Movies
  - Concept
  - Comparison with other media
  - Configuration of the first system
    - Software
    - Hardware
  - Evaluation and problems
- Description of the Second System
  - Improvement
  - Software system structure
  - Hardware system structure
- Example of Interactive Story Production
  - An interactive story
  - Interaction
- Conclusions
- Bibliography

Introduction

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

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

Based on this viewpoint, we have been conducting research on interactive movie production by applying interaction technology to conventional movie making techniques. As an initial step in creating a new type of movie, we have produced a prototype system[4]. Based on this system, we are currently developing a second prototype system with many improvements. This paper briefly describes the first prototype system and outlines its problem areas and required improvements. The paper also introduces the configuration of the second prototype system, which is now under development by incorporating the described improvements.

<-- Table of Contents

Outline of Interactive Movies

Concept

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

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

Interactive movies have the following functions:

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

Comparison with other media

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

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

(2) Movies
Movies use dynamic images and sound to provide strong input to the human visual and auditory sensory systems and draw audiences into a cyberspace. The rapid advance of CG technology in recent years has enabled the production of very realistic images of extraordinary worlds and phenomena that do not exist or happen in the real world. There have been attempts to add an interactive function to those movies. However, such efforts were limited to a primitive level: several different story developments were prepared, and an audience selected from available options.

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

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

Configuration of the first prototype system

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

'Software'

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

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

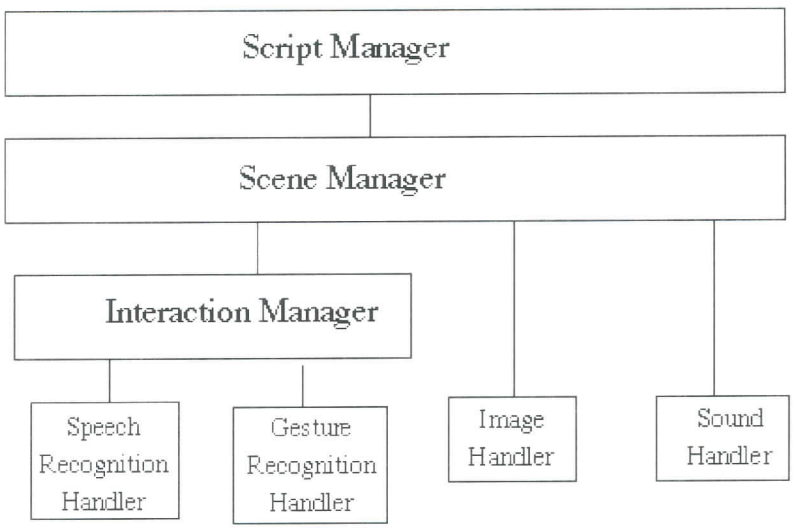


Figure 1. Software configuration of the first prototype system.

'Hardware'

Figure 2 shows the hardware configuration. The system consists of an image output subsystem, a speech recognition subsystem, a gesture recognition subsystem and a sound output subsystem. In the image output subsystem, a high-speed CG-generating workstation is used for the visual image output. Character images and background CG images are stored as CG animation data and used for the real-time generation of CG images. The speech recognition operations are executed by a single workstation. The gesture recognition function is also executed by a workstation. The sound output subsystem consists of several PCs. Background music, sound effects and character dialogs are produced simultaneously by this subsystem.

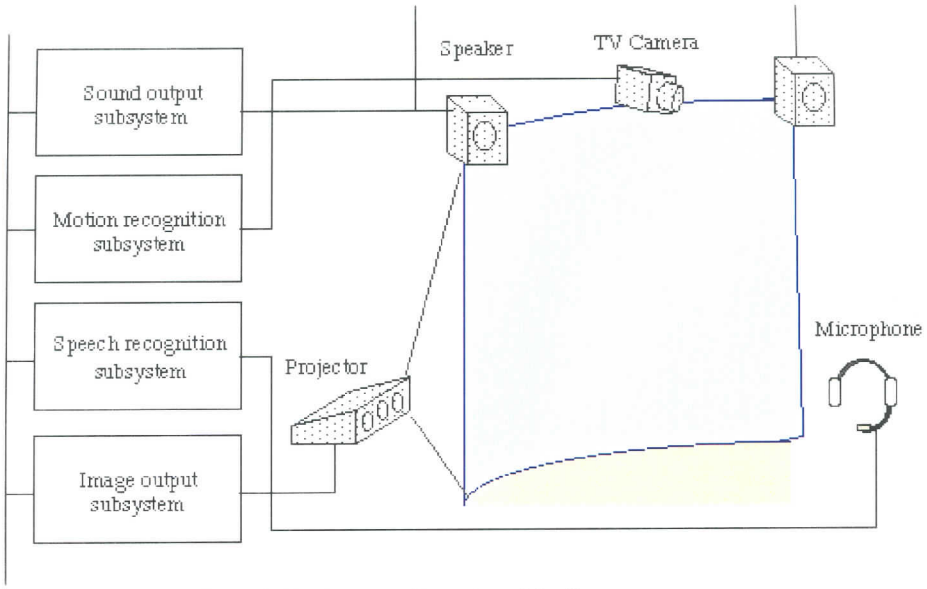


Figure 2. Hardware configuration of the first prototype system.

Evaluation and problems

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

(1) Participation in cyberspace

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

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

(2) Interaction

a) Frequency of interaction
Interaction in the first system was generally limited to change points in the story, so the story progressed linearly along a predetermined course like a movie except at these change points. Even though there are certain advantages to this technique, such as being able to use movie story development technology and expertise, the disadvantage of fixed story elements created as in a movie is that the player seems to end up a spectator who finds it difficult to participate interactively at points where interaction is clearly required. The limited opportunities for interaction create other drawbacks for the player, such as having little to distinguish the experience from watching a movie and having a very limited sense of involvement.

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

<-- Table of Contents

Description of the Second System

Improvement

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

(1) Participation in cyberspace

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

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

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

System for controlling the avatar: The basic control system inputs player movement by using magnetic sensors and uses that movement to map avatar movement in a motion capture system. Giving autonomy to avatar movement enables complex movement that combines autonomous avatar movement with player movement. By varying the proportion of each movement with time and circumstance, player movement can be used directly or the player can introduce desired movement that together add diversity and depth to the relationship of the player to cyberspace.

(2) Interaction

a) Introduction of interaction any time

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

b) Introduction of multimodal interaction
The following interactive functions were added to the primary interactive function of voice recognition. Emotion recognition: To realize interaction at any time, an emotion recognition capability was introduced. When players utter spontaneous utterances, the characters react by using their own utterances and animations according to the emotion recognition result. Emotion recognition is achieved by using a neural-network-based algorithm[9].

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

Gesture recognition: We captured motion with magnetic sensors and used an HMM to process data from the sensors in order to recognize 3-D gestures, minute gestures and gestures under low-light conditions. Gesture recognition results are used for SCI. Fig. 3 shows an overview of the second system equipped with these functions.

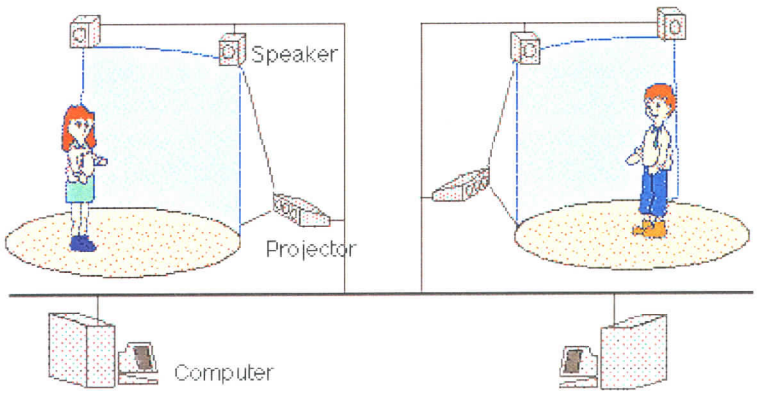


Figure 3. Appearance of the second system.

Software sytem structure

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

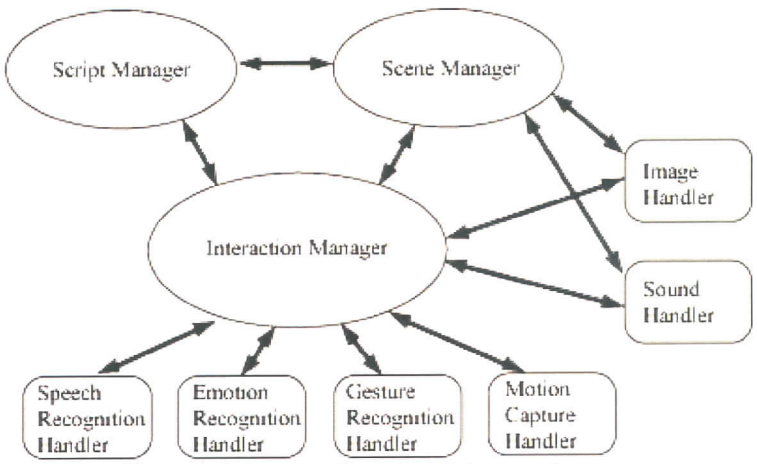


Figure 4. Software configuration of the second system.

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

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

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

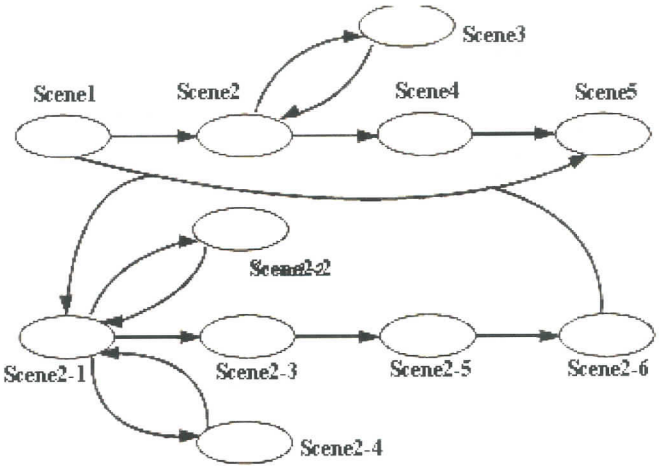


Figure 5. Sene transistion network.

(3) 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 derive from the following elements.

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

The script for each scene is stored ahead of time in an event network, and the scene manager generates each scene based on data from the script manager via a script in the format shown in Fig. 6. The timing for transition from one event to the next was controlled by the scene manager in the first system, but absolute time cannot be controlled in the second system because it incorporates the concept of interaction at any time. However, relative time or time order can be controlled in the second system, so the action selection network was applied here as well. The following describes how this works.

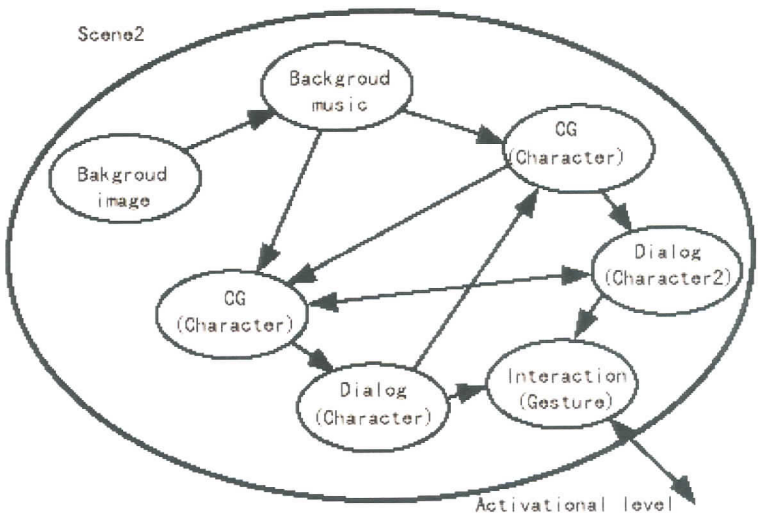


Figure 6. An example of scene data.

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

(4) Interaction manager
The interaction manager is the most critical aspect for achieving anytime interaction. Figure 7 shows the structure of the interaction manager. The basis for anytime interaction is a structure where each character (including the player's avatar) is allotted an emotional state, and interaction input from the player as well as interaction between the characters determines the emotional state as well as the response to that emotional state for each character. Some leeway is given to how a response is expressed depending on the character's personality and circumstances. The interaction manager is designed based on the concepts outlined below.

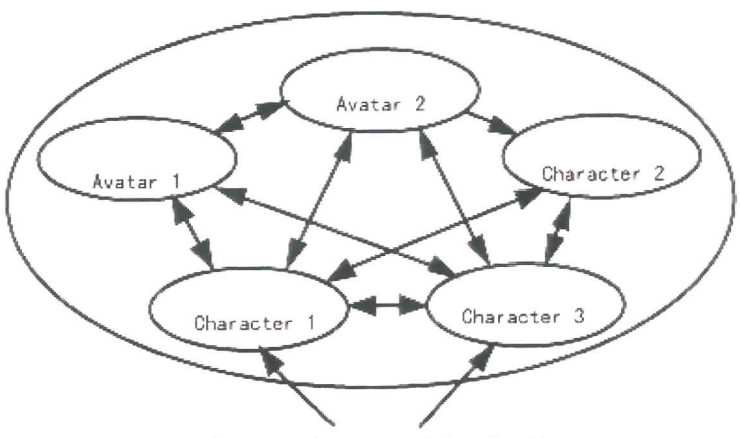


Figure 7. Structure of interaction manager.

1) Defining an emotional state
The state and intensity of a player's (i = 1, 2...) emotion at time T is defined as follows:

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 follows:

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

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

{Ep(i, T)} -> {Eo(j, T + 1)}

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

sp(i, T) -> 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) = · sp(i, j, T)

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

if so(i, T) > THi
then Eo(i, T) -> Ao(i, T), Eo(i, T) -> so(i, j, T)
so(j, T + 1) = · 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.

Hardware sytem structure

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

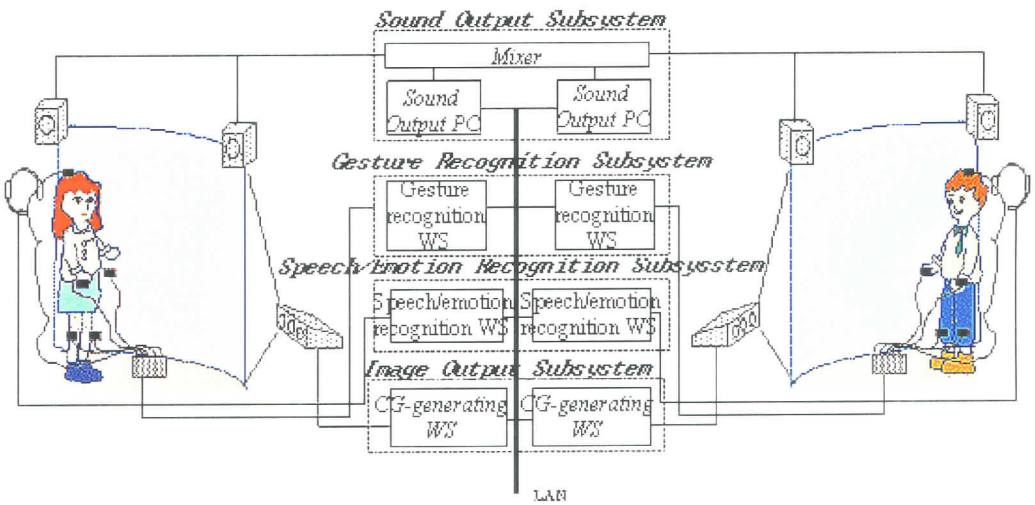
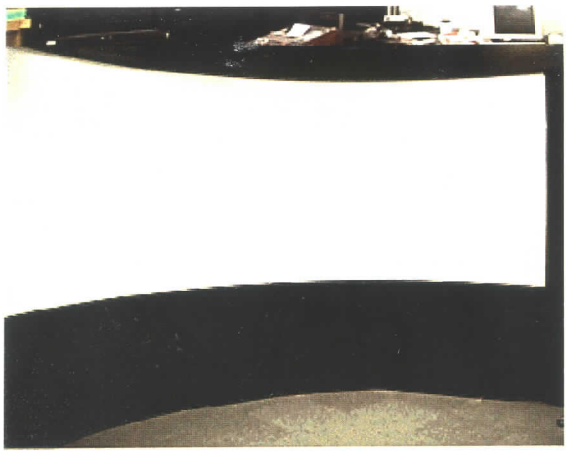


Figure 8. Hardware configuration of the second system.

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

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



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

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

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

<-- Table of Contents

Example of Interactive Story Production

An interactive story

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

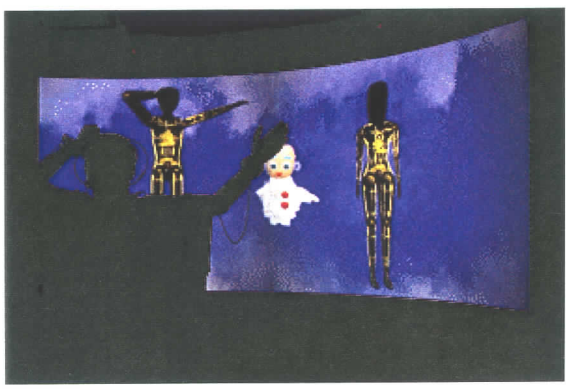
a) There are two main characters, Romeo and Juliet, in the story and, therefore, this story supplies a good example of multi-person participation.

b) "Romeo and Juliet" is a very well known story, and people have a strong desire to act out the role of hero or 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 their souls are sent to Hades, where they find that they have totally lost their memory. Then they start their journey to rediscover who they are and what their relationship is. With various kinds of experiences and with the help and guidance of characters in Hades, they gradually find themselves again and finally go back to the real world.

Interaction

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



Romeo controls his avatar.

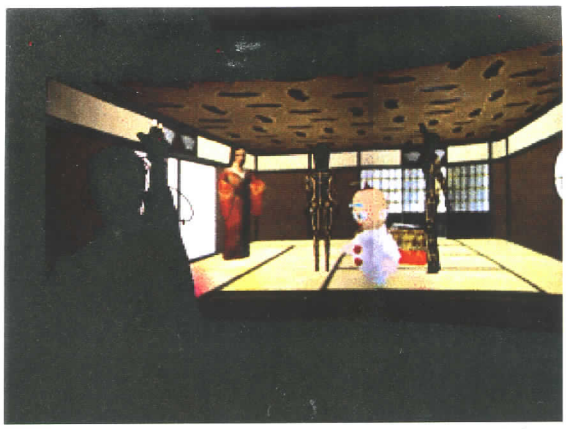


Figure 10. Examples of interactions between participants and the system.

Download QuickTime Movie (22sec. 1MB)

<-- Table of Contents

Conclusions

Interactive movies are a new type of media that integrates various media types including telecommunications, movies, and video games. In interactive movies, people enter cyberspace and enjoy the development of a story there by interacting with the characters in the story. In this paper, we first explained the concept of interactive movies and briefly explained our first prototype system. Based on an evaluation of this system, we identified several problems in the system that needed to be improved. One is the lack of frequent interactions and the other is single-person participation. To improve these deficiencies, we are developing a second system. We explained two significant improvements incorporated into the second system: interaction at any time and two-person participation through a network. We described the software and hardware configuration of the second system while putting emphasis on these improvements. We also described two interaction technologies that are introduced for the first time in the second system and showed that are the keys to realizing these improvements. Finally, we illustrated the operation of our Second System with the example of our interactive production of "Romeo and Juliet".

<-- Table of Contents

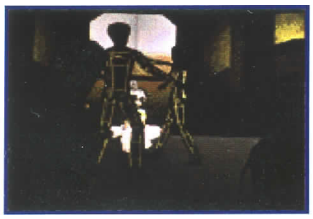
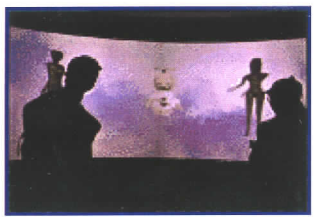
Bibliography

[1] Ceram, C. W.: Eine Archaologie des Kinos. Rowohlt Verlag, Hamburg, 1965.
[2] Maes, P., et al.: The ALIVE system: Full-body Interaction with Autonomous Agents. in Proceedings of the Computer Animation 95 Conference, 1995.
[3] Maes, P. : How to do the right thing. Connection Science, Vol.1, No.3, 291-323, 1989.
[4] Nakatsu, R., and Tosa, N.: Toward the Realization of Interactive movies - Inter Communication Theater: Concept and System. in Proceedings of the International Conference on Multimedia Computing and Systems 97, 71-77, 1997.
[5] Noma, H., et al. : Multi-Point Virtual Space Teleconference System. IEICE Trans. Commun., Vol. E78-B, No. 7, 1996.
[6] Ohya, J., et al. : Virtual Kabuki Theater: Towards the Realization of Human Metamorphosis Systems. in Proceedings of 5th IEEE Workshop on Robot and Human Communication (RO-MAN 96), 416-421, 1996.
[7] Shimizu, T., et al.: Spontaneous Dialogue Speech Recognition Using Cross-Word Context Constrained Word Graph. in Proceedings of ICASSP 96, Vol. 1, 145-148, 1996.
[8] Sommerer, C., and Mignonneau, L.: A-Volve: An Evolutionary Artificial Life Environment. in Proceedings of the Fifth International Workshop on the Synthesis and Simulation of Living Systems, 167-175, 1997.
[9] Tosa, N., and Nakatsu, R. : Life-like Communication Agent - Emotion Sensing Character MIC and Feeling Session Character MUSE -. in Proceedings of the International Conference on Multi-media Computing and Systems, 12-19, 1996.
[10] Wren, C. R., et al.: Pfinder: Real-Time Tracking of the Human Body. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, 780-785, 1997.

<-- Table of Contents

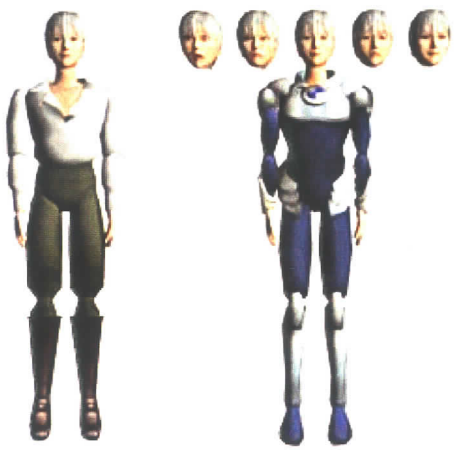
RETURN
NAOKO TOSA tosa@mic.atr.co.jp

Interactions




HOME
NAOKO TOSA tosa@mic.atr.co.jp



HOME