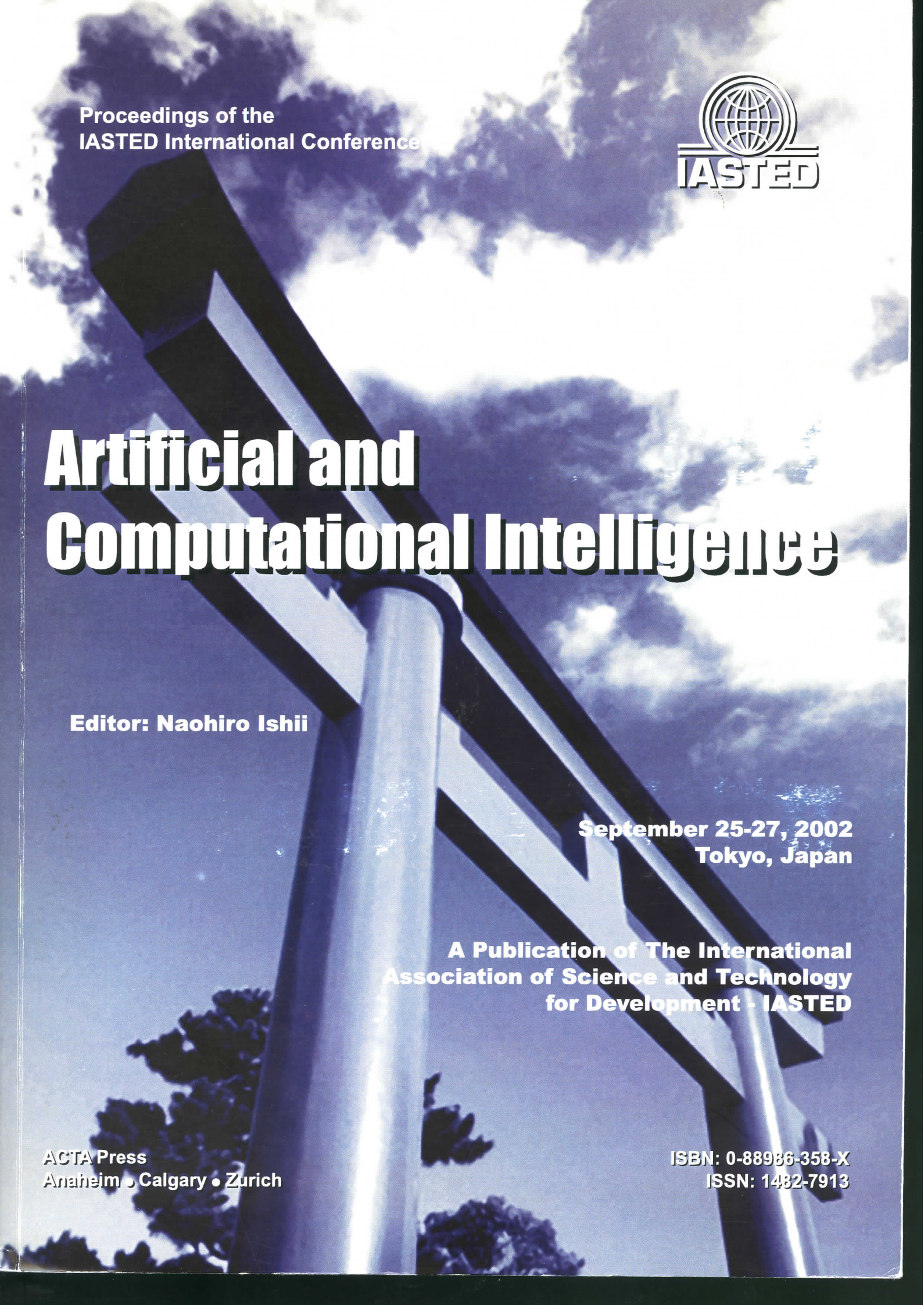

Proceedings of the
IASTED International Conference

IASTED

Artificial and Computational Intelligence

Editor: Naohiro Ishii

September 25-27, 2002
Tokyo, Japan

A Publication of The International
Association of Science and Technology
for Development - IASTED

ACTA Press
Anaheim • Calgary • Zurich

ISBN: 0-88986-358-X
ISSN: 1482-7913

Proceedings of the IASTED International Conference
ARTIFICIAL AND COMPUTATIONAL INTELLIGENCE
September 25-27, 2002, Tokyo, Japan

INTERACTIVE COMEDY: LAUGHTER AS THE NEXT INTELLIGENCE SYSTEM

Naoko Tosa, Ryohei Nakatsu
Center for Advanced Visual Studies, Massachusetts Institute of Technology
265 Massachusetts Avenue, Cambridge, MA 02139 USA
naoko@mit.edu
Interaction & Intelligence project (JST) ATR HIS Labs
2-2-2 Hikaridai Keihanna Science City
Kyoto 619-0288 Japan

Abstract

Generally, technology looks at the external value of an object. At a very deep level, however, laughter comes from the touch of living—decidedly an internal value. It is very difficult to fully comprehend this deep relationship. Our personal feeling is the key to grasping the phenomenon of laughter, which is not an issue of analysis or understanding but one of the mind's spontaneity, such as the emergence of an idea or a burst of imagination. Laughter is an intelligence that emerges through sympathetic communications and, therefore, has great power.

Recently, artificial intelligence research has focused too much on analysis and knowledge and, consequently, risks being capable of achieving only boring results. Genuine intelligence is not boring, so the intelligence we develop should not be boring. To meet this challenge, I have developed a comedy system that enables humans and computers to interact and create laughter.

Key Words:

Emotional Intelligence, Computing for information of Culture database, creative Intelligence Communication system

1. Introduction

Laughter is an essential element of verbal conversation, and we transmit information by verbal conversation. At the same time, verbal conversation transmits our emotions and basic desires [1]. Therefore, verbal conversation is the basis for maintaining human relations and a social life. In this type of conversation, such factors as voice tone, timing, and speed play very important roles [2]. By controlling them, we enjoy our conversations, sustain human relations, and energize our social life [3].

2. Laughter as Emotional Intelligence

Laughter has a great power and an important impact on our body. Laughter constantly moves back and forth, up and down, inside and outside of our world.
It can elevate our minds and banish deceptive gloss. Laughter that comes from the deepest crevice of our mind can break the wall of our anxiety and loneliness. Its effect can match that of tears. Our laughter can also make a strong impact on the minds of those around us.

3. Comedy as good communication

Communications has the power to entertain, and this power of communications has a strong appeal to us. Therefore, an artfully conducted conversation entertains us. In the entertainment discipline of comical dialogues, conversation has been refined to a level of sophisticated entertainment. In Japan, we have our own form of comical dialogue called Manzai. We enjoy Manzai by listening to the dialogue of two Manzai comedians. The two comedians play the role of "Boke" and "Tsukkomi." Boke plays the role of an ordinary person who expresses his/her everyday opinions. Sometimes these opinions are superficial and boring, far from nuanced reality. Tsukkomi plays the role of a sensitive person who is keen to detect these opinions and becomes offensive to Boke. Boke usually shrinks from the opinions of Tsukkomi but sometimes awakens to reality. For the conversation between Boke and Tsukkomi, such factors as voice tone, timing, and speech are as critical as the conversation's content. When these factors are skillfully controlled, the audience is fascinated and becomes involved in the conversation. They want to become Manzai comedians themselves.

Figure 1. Japanese Comedy: Manzai

4. An Interactive Comedy System

An interactive comedy system is designed to realize the passive audience member's dream of becoming a Manzai artist. You play the role of defence and the computer plays the role of offence. When you speak, the computer analyses the content, emotion, speed and timing of your utterance and utters back an appropriate offensive phrase. Then you speak back. By interacting with this system, you feel the joy of conversation itself. At the same time, you subconsciously learn how to effectively control timing, utterance speed, and emotional expression.

Figure 2. Interactive Comedy system

5. Emotion model of Computer Comedian

The Computer Comedian can recognize eight emotions from the intonation in the voice and respond to these emotions. For example, unless someone speaks, the Computer Comedian sleeps; when someone speaks, the Computer Comedian responds with "What's shakin'?" when in a good mood or "Oh, I'm sorry, were you talking to me?" when in a bad mood. When made fun of in a low voice or when teased, the Computer Comedian becomes angry. When someone whistles, it becomes excited and jumps, and when a person makes an unflattering face and coughs, it becomes sad, covers its face with its hands, and looks backward. Occasionally, it complains to a person that shows disgust. The emotions model of Computer Comedian is shown in the 2D coordinates in Fig. 3. The X axis represents goodness/badness and the Y axis represents the strength of the emotion.

The eight concrete emotions and the characteristics (denoted by *) of the input speech corresponding to them are as follows.

(1) Joy (happiness, satisfaction, easy comfort, smile)
*High and strong voice, whistling, etc.
(2) Anger (anger, bad mood)
*Scolding in a low and strong voice.
(3) Surprise (wonderment, shock, surprise)
*Scolding in a sudden and loud voice.
(4) Sadness (pity, tears, loneliness)
*Low and weak voice, coughs, etc.
(5) Disgust - - - * Tired voice with disgust.
(6) Teasing - - - * Teasing in a high and light voice.
(7) Fear - - - * Fearful voice with a flat and low voice.
(8) Neutral - - - * Natural voice.

An example from the Database for Interactive Comedy is shown below.

People = (P) Computer = (C)
(P) Well, the nice weather is finally here.
(C) It's about time.
(P) I can't wait to get to the beach.
(C) Are you crazy, the water here is freezing.
(P) It's pretty amazing speaking with a "Manzai" Computer.
(C) Stop sucking up!
(P) If that's the way you're going to be, then maybe I should leave.
(C) Chill out.
(P) I'm glad you found time to join us today despite your busy schedule.
(C) You make me feel so good.
(P) Maybe you are not that busy, seeing as you made it here today.
(C) What is that suppose to mean?
(P) Can you guess how many years we have been carrying out this research? 300years!
(C) You lie like a dog.

Emotion recognition is carried out using human voice intonation information and an artificial neural network [4].

An example of emotional dialogue is shown below.

<Excited>
Awesome!
Ooh, you're giving me goose bumps!
<Neutral>
Oh, I'm sorry, were you talking to me?
Tell me more.
<Anger>
Keep your pants on.
Bite me.
<Disgust>
Take it easy there, cowboy.
I know your lips are moving, but all I hear is blah blah, blah.
<Teasing>
Talk to the hand.
Are you trying to seduce me?
<Sadness>
Don't make me cry.
Maybe you're right. Maybe I should just go.
<Surprise>
I think you're lying to me....
<Greeting>
What's shakin'?
<Fear>
You're scarin' me now
Don't go psycho on me.

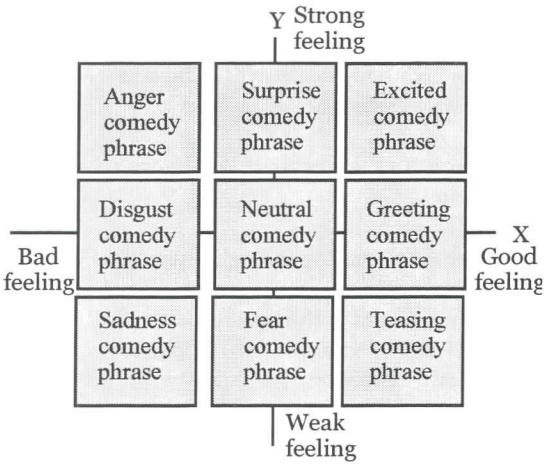

Figure 3. Comedy emotion model

6. How does it work?

6.1 Components of the system's architecture

The system consists of three parts. In the content analysis part, the meaning of the input speech is recognized using speech recognition technology. In the emotion analysis part, the tone of the input speech is analyzed.[5] The speed and the timing of the input speech are also analyzed. By combing these two results, the level of user involvement, in other words how deeply a user is involved in the conversation, is decided. In the character synthesis part, using the content analysis and the involvement analysis results, the speech output and the facial animation of the comedy character is generated. Accordingly, the user enjoys a laugh-filled conversation with the system

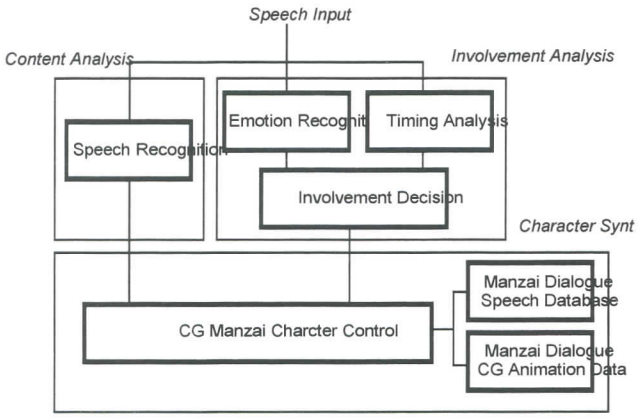

Figure 4. System Configuration

6.2 Processing

Emotion Recognition[4]

1) User voice input is first made by microphone. Then speech processing is used to achieve spectrum, pitch and tone.

2) Emotion recognition result is generated from the learned neural network by a human emotional voice. Emotion recognition result output is transformed to a two-dimension coordinate and then sent to the script manager.

Next, the system carries out adaptation to the voice of a new user by shifting the point in the 2D emotion plane corresponding to the recognition of the neutral voice to (0, 0).

Voice recognition

The user's voice is detected from microphone analysis by the voice recognition system. If the system recognizes a comedy-registered dictionary designed for the voice recognition system, it sends a relevant comical reply to the script manager.

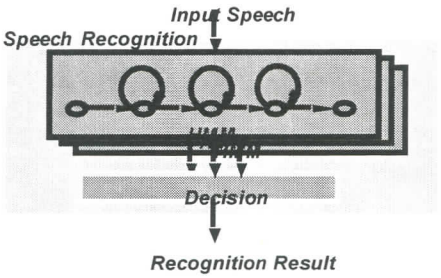

Figure 5. Speech analysis diagram

Script Manager

The script manager selects a facial expression from the computer graphics data and outputs voice data from the result of emotion recognition and voice recognition. The system carries out emotion recognition and speech recognition simultaneously. If the speech recognition result is rejected because it does not have a high enough recognition score, the system uses the emotion recognition result.

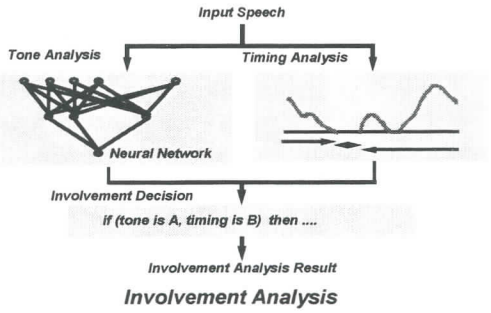

Figure 6. Involvement analysis diagram

Real-time Computer Graphics generation [6]

The Computer Graphics Character displays facial animation from the data sent from the script manager. At the same time, the system makes the output sound of narration.

The system making the emotional expression depends on several types of emotion data represented by facial expressions and comical emotions.

Output comedian phrase

The script manager finally decides whether the speech recognition result or the emotion recognition result should be used by the system.

1) If the result of speech recognition is adopted, the corresponding output phrase is selected from the database and sent to the user both as a response voice and as a character sequence on the display.

2) If the emotion recognition result is adopted by the system, wave file data stored according to each point of the 2D emotional plane are selected and sent to the user. Multiple wave file data can be stored on the same emotion plane.

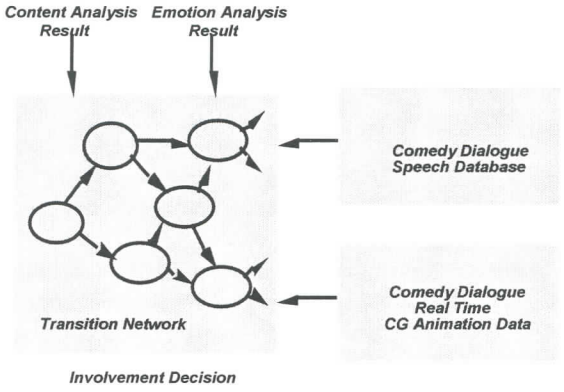

Figure 7. Real-time based Computer Graphics Character Synthesis

7. Conclusions

This system was demonstrated and broadcast on "Most Interesting News Digest of 2001," the special New Year's Eve program of Asahi TV. We received the Grand Prize in the Future Robot category. This system is simple, but the technology introduced here to make people laugh has a strong emotional resonance that can spread worldwide. Laughter opens the mind and stimulates real communication among us. Computers can actually make us laugh, and this is very intelligent interaction. However, when people are too self-conscious in trying to have fun, communications do not go smoothly and become difficult. This means that laughter is related to the nonverbal and subconscious aspect of our intelligence. If this type of technology were introduced into a computer system or robot in the future, the tendencies toward hospitality and caring for others would become more important in our society. Also, a comedy's script is a very important tool to make us laugh. Furthermore, laughter is strongly dependent on local culture, and in each country the meaning of laughter is very different. For the Japanese version of Interactive Manzai, the script was developed by the Yoshimoto Entertainment Company in Japan. Then when I showed Interactive Comedy in Boston, Improv Asylum, which is one of the best comedy groups in the city, translated the dialogs into an American style. This spring, we performed the English version of Interactive Comedy System at the Japan Pop Culture Festival, one of the events of the Haru Fest in Boston, at the MIT Media Labs Bartos Theater. This perfomance was organized by the Consulate General of Japan in Boston. In evaluating the system's effect, we must bear in mind that comedy is very different depending on the area and culture. A comedy system also has to contend with the same features of user reaction as does a conventional comedian. In the Osaka area, where people use daily conversation for comedy, there was great excitement about our interactive comedy system. On the other hand, Tokyo people reacted with a sneer when viewing interactive comedy. However, when people start the interaction, good communication starts flowing because people are laughing. In Boston, I first showed the system by directly translating from Japanese to English and, not surprisingly, the Americans did not laugh. When the Boston comedy group translated the material into a more American style of comedy and performed it at the MIT Theatre, people laughed at each interaction of the Interactive Comedy System. Consequently, we also performed cultural computing.

8. Acknowledgements

The authors express their thanks to the contributions of the legendary Japanese entertainment company Yoshimoto Kogyo Co., Ltd and Boston's comedy group "Improv Asylum," for creating the comedy script for us. We also thank the Consulate General of Japan in Boston for their arranging the exhibition of this system.

References

[1] Pinhanez, Claudio and Bobick, Aaron. " Using Computer Vision to Control a Reactive Computer Graphics Character in a Theater Play." Proceedings of ICVS'99, Las Canarias, Spain, January 1999.

[2] Picard, R. W. and Healey, J. "Affective Wearables" Personal Technologies, Vol. 1, No. 4, pp. 231-240, December 1997.

[3] Blumberg, Bruce and Galyean, T. " Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments. " Computer Graphics, SIGGRAPH '95 Proceedings, Los Angeles, California, pp. 47-54. August 1995

[4] Tosa, Naoko and Nakatsu, Ryohei, "Life-Like Communication Agent—Emotion Sensing Character "MIC" & Feeling Session Character "Muse,"" Proceedings of the third IEEE ICMCS, pp. 12-19 (1996)

[5] Nakatsu, R., Tosa, N., Ochi, T., and Suzuki, H., "Concept of Interactive Movie Systems and Construction Example," Electronic Information Communications Society Papers, DII, Vol. J81-D-II, No. 5, 1998, 944-953

[6] Tosa Naoko, "The Esthetics of Recreating Ourselves " SIGGRAPH □ 96 Cours Note No. 25, " Life-like Believable Communication Agents" (1996).