

# IMAGE/SPEECH PROCESSING THAT ADOPTS AN ARTISTIC APPROACH - TOWARD INTEGRATION OF ART AND TECHNOLOGY -

*Ryohei Nakatsu*

ATR Media Integration & Communications Research Laboratories

2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

nakatsu@mic.atr.co.jp

## ABSTRACT

In the areas of image/speech processing, researchers have long dreamed of producing computer agents that can communicate with people in a human-like way. Although the non-verbal aspects of communications, such as emotions-based communications, play very important roles in our daily lives, most research so far has concentrated on the verbal aspects of communications and has neglected the nonverbal aspects. To achieve human-like agents we have adopted a two-way approach.

1. To provide agents with nonverbal communications capability, engineers have started research on emotions recognition and facial expressions recognition. 2. Artists have begun to design and generate the reactions and behaviors of agents, to fill the gap between real human behaviors and those of computer agents.

## 1. INTRODUCTION

In this paper, the possibilities that might emerge by combining image/speech processing technologies and art are discussed. Generally, engineering technologies tend to open at the forefront but eventually tend to dissociate from human factors in the name of "high tech". In contrast, art expresses the deepest parts of humans such as emotions or senses. Putting things in perspective, art and technology seem to be like oil and water.

In ancient times, however, a lot of people were both an engineer and an artist. In modern times, on the other hand, the gap between art and technology has been getting wider and wider with the rapid progress in science and its specialization.

In the field of communications, the development of new communication systems and services for the next century is expected by the utilization of the high technology called "Multimedia". However, we cannot deny that there is some anxiety that our future society, to be full of high-technology equipment, will lack human compassion and therefore, will be gloomy. The reason for this is that recent technologies have been advancing in a direction that ignores the human senses and emotions.

Yet, we think it is important to develop services and systems while considering the human senses and emotions. For this reason, we believe it is necessary for engineers to work together with people who can handle these human factors, such as artists. Based on this point of view, in our research laboratories, we are carrying out research, aimed at new communications technologies, based on collaboration between artists and engineers. In this paper, basic

concepts and examples of such trials will be stated.

## 2. COMMUNICATIONS AND IMAGE/SPEECH PROCESSING

### 2.1 Intellectual activities of human beings and image/speech processing

Handling the intellectual activities of human beings is Artificial Intelligence (AI) 's main subject. Among the various kinds of intellectual activities we focus on the functions of human communications. The main reason is that general aspects of intellectual activities are expressed in communications. In this area, so far, engineers have been concentrating their research on robots or computer agents that have functions to communicate with human beings. The major part of this research, sadly to say, has been emphasizing only the verbal aspects of communications. For example, speech recognition has been aiming at extracting basic meanings, that is, verbal information. However, it has been recognized that in daily life, the transfer of emotions and senses, that is, non-verbal communications, also plays an important role. For example, speech includes speaker-related information and emotions-related information in addition to verbal information. In speech recognition, however, non-verbal information has been ignored and treated as noise.

Creating human-like computer agents or characters requires the research and development of technologies concerned with non-verbal communications. Agents adopting such technologies may be able to have hearty communications with human beings.

### 2.2 Communication model

Figure 1 shows a human communication model. It should be noted that this model has a similar construction to the human brain. In the outer layer, which corresponds to the new cortex in the brain, there is a layer that controls communications based on the use of a language. Researchers in the AI field have been studying the mechanism of this layer.

Speech recognition is a typical example of AI research. In the field of speech recognition, research has been done for many years on algorithms that can achieve a high recognition performance by handling only the logical information included in speech. As stated above, logical information is only a part of the whole information that constitutes speech. Other rich information, like information on emotions or senses are also included. Such information are

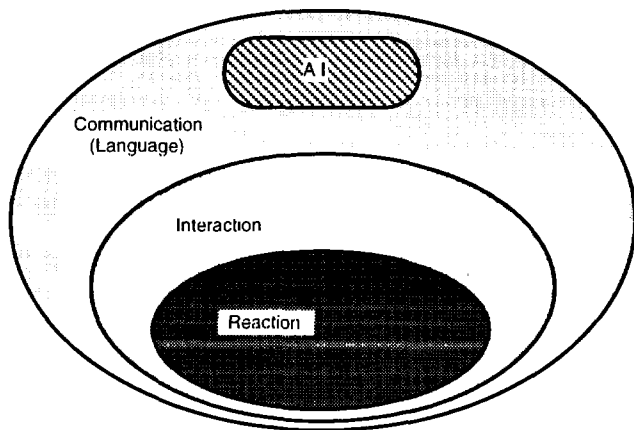


Fig.1 Communication Model

considered to be created at the deeper level layers, that is, the Interaction and Reaction layers as indicated in Fig. 1.

The interaction layer controls actions to maintain the communication channels, like nodding, controlling the speech rhythms, or managing the changes of speech turn. This layer plays an important role in achieving smooth human communications. Under this layer is the Reaction layer, which controls more basic actions. Examples of these actions include turning one's face toward the direction from which a sound had come or closing one's eyes upon suddenly sensing a strong light. Such functions were obtained in ancient times when human beings were still uncivilized.

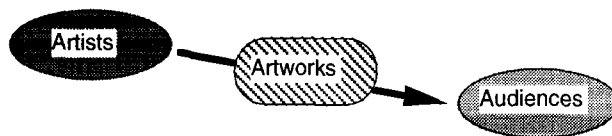
Thus, not only the handling of logical actions and information but also the handling of functions at deeper layers plays an important role in human communications. These functions create and understand non-verbal information like emotions and senses. The reason why efficiency in speech recognition has so far been limited is because such essential information has been neglected as noise. Therefore, in order to understand general human communications functions including the sending-receiving of emotional information other than logical information, it is necessary to research the mechanism of the Interaction layer and Reaction layer and to integrate the results with the functions of the Communication layer. By doing so, agents with human-like behaviors can be created.

### 3 APPROACH AIMING AT THE INTEGRATION OF ART AND TECHNOLOGY

In the previous section, the necessity of studying the action mechanisms of the deeper level layers in human communications was explained. This section proposes the idea of integrating technology and art.

As stated before, in the engineering field, research is being done targeting the handling of logical information in human communications. As the research advances, however, it is becoming clear that the mechanisms of deeper level communications, like communications based on emotions or senses do play an essential role in our daily communications. It is, therefore, inevitable to be able to handle information on emotions and senses, which had not been handled in the engineering field up to now. On the other hand, artists have long handled human emotions and senses. Therefore, further development is expected by having engineers collaborate with artists.

#### Traditional Art



#### Interactive Art



Fig.2 From Traditional Art to Interactive Art

Art too has seen a notable movement recently. This is due to the emergence of a field called Interactive Art. The important function of art is to have an artist transfer his/her concepts or messages to an audience by touching their emotions or senses. In the long history of art, this means of communications has been refined and made sophisticated. However, it cannot be denied that in traditional art, the flow of information in communications has been one-way, that is, information is transferred from the artist to a passive audience.

With Interactive Art, the audience can change expressions in art works by interacting with them. That is, the audience provides feedback to the various art works and this consequently enables information to flow from the audience to the artist. Therefore, in Interactive Art, information flow is both ways, that is, true communications is achieved. A comparison of information flows between traditional art and Interactive Art is illustrated in Fig. 2.

At the same time it should be pointed out that this Interactive Art is still developing and that interactions remain at the primitive level, like causing a change by pushing a button. Therefore, it is necessary for Interactive Art to adopt image/speech processing technologies to raise primitive interactions to the communications level.

For this aim, from an engineering viewpoint, collaboration with art is required to give computers human-like communications functions. From the art side, adopting new technologies is necessary to improve the current Interactive Art, from the level of interactions to that of communications. As both approaches share the same target, the time is ripe for collaboration between art and technology to progress.

### 4. EXAMPLE OF APPROACHES INTEGRATING ART AND TECHNOLOGY

In our laboratory, based on the above idea, we started to employ artists in the Interactive Art field from last year and began new attempts to carry out research based on collaboration and joint activities between artists and engineers. In the following, some examples of research activities in our laboratory are described.

#### 4.1 Emotional agent "MIC"[1][2]

In human communications using voices, emotions play a very important role. Sometimes, information on emotions is more essential than the logical information included in speech. This can be confirmed from the fact that babies start to recognize emotional

information before they can recognize information in their mothers' voice. In the case of adults too, we can recognize what other people want to say at a deeper level by integrating information on meanings and emotions included in speech. This is the key for communications to proceed smoothly. Unfortunately we have to say that, in the field of AI so far, focus has been on recognition of only meaning information and emotions have been neglected as noise. In order to create an agent with human-like behaviors, therefore, it is necessary to add functions enabling it to recognize emotions and to react to them.

"Neuro Baby" [3] was produced according to this idea. Neuro Baby is a computer character that is capable of recognizing four emotions included in speech and reacting to them by changing his facial expressions. Based on the experiences in developing and exhibiting Neuro Baby, we then produced "MIC" last year. In comparison with Neuro Baby, MIC has the following improvements.

(1) A better ability in non-verbal communications : MIC is a character that reacts to emotions involved in speech just like Neuro Baby. He can recognize eight emotions (joy, anger, surprise, sadness, disgust, teasing, fear, and normality) to Neuro Baby's four. The reaction patterns of the character were also improved : A CG image of a full-length portrait was created and the character's emotional reactions were expressed by whole body reactions as well as facial expressions.

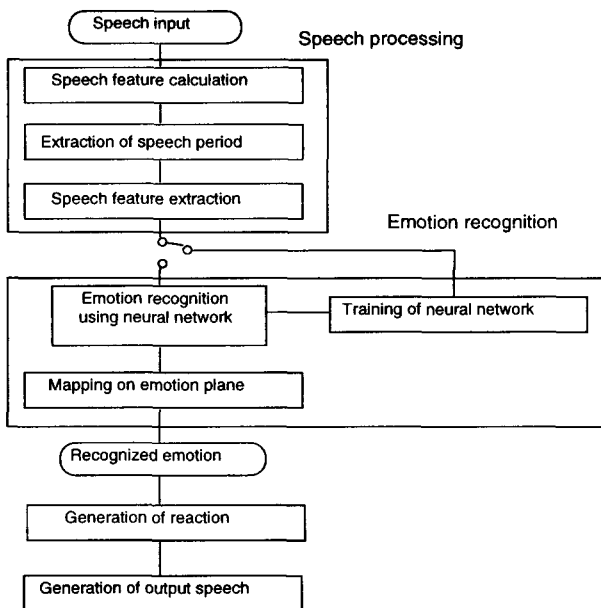


Fig. 3 Block Diagram of the Processing Flow

(2) Improvement of functions for emotions recognition : By adopting the following methods and technologies, functions for emotions recognition were improved.

- To improve the emotions recognition capability of MIC, a combined-type neural network architecture was introduced. Eight neural networks that corresponded to each of the eight emotions were prepared and emotions recognition was achieved by feeding feature parameters into these eight networks in parallel.
- By using speech data, that is, various kinds of phonetically

balanced words uttered by many speakers, as training data, speaker-independent and content-independent emotions recognition became possible.

The whole processing flow is shown in Fig. 3 and the construction of the emotions recognition part is presented in Fig. 4. According to the emotions recognition results, MIC reacts by changing his facial expressions and body actions. These reactions were carefully designed and developed by an artist. This approach, combined with emotions recognition technology, has enabled the agents to behave truly human-like. An example of MIC's representative reaction patterns is given in Fig. 5.

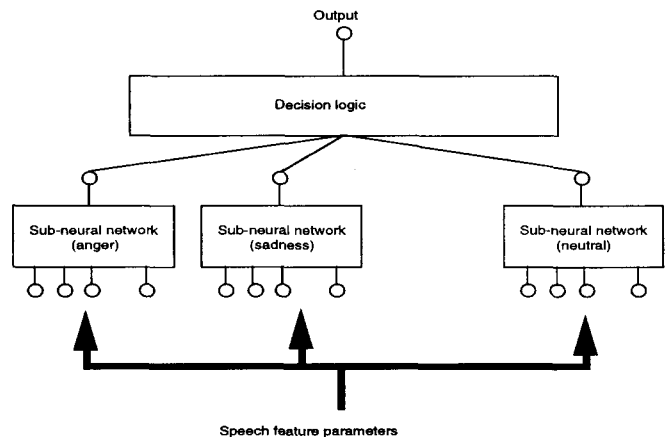


Fig. 4 Configuration of Emotions Recognition Part



Fig. 5 An Example of MIC's Emotional Expression

#### 4.2. Virtual KABUKI [4][5]

Facial expressions play an important role in human natural communications. We communicate with other persons smoothly by recognizing their emotions through their facial expressions and also by expressing our emotions through our facial expressions. In order to design an agent with a human shape in a virtual space, therefore, a technique is required to extract facial expressions from its image in real time and to reproduce them with a three-dimensional face model. For this objective, we studied the real-

time recognition of facial expressions and reproduction technologies. As an example of applying this technique to the creation of a human-like agent, we examined the reproduction of a three-dimensional face model of a KABUKI actor.

The flowchart for recognition of facial expressions and creation processing is shown in Fig. 6. The recognition of facial expressions and creation processing system consist of three parts : extraction of expressions, face reconstruction, and face modeling. The face model must be created beforehand ; a three-dimensional model of the face is created in the form of a wire frame model. With this wire frame model, the facial shape is made similar to an assembly of small triangular patches and the color texture of the face is rendered on these triangle patches. In order to extract facial expressions in real time, the subject puts on a helmet and a small video camera attached to the helmet takes images of the subject's face. If the position or direction of the head changes, the helmet follows these changes and always extracts facial images stably. Next, DCT conversion (discrete cosine conversion) is carried out on the obtained images by the camera and changes of facial compositions, such as the eyes or mouth opening/closing, are extracted. Information on the changes are reflected in the transformation of the three-dimensional model and the extracted facial expressions are reconstructed as facial expressions of a KABUKI actor. An example of a reconstructed KABUKI actor is shown in Fig. 7.

Of course, a key point of this system is, technically, the extraction of facial expressions in real time. At the same time, the transformation technique, which extracts the expressions and reproduces them as a KABUKI actor's expressions, is essential. Note that an artist creates the KABUKI actor's face model to add an artistic touch. By adding such an artistic element, anyone can transform himself into a KABUKI actor.

By extending this idea, the possibility of new entertainment, such as the combination of a role playing game and a movie, where someone enters into a virtual space and experiences various stories, can be born.

## 5. CONCLUSION

In this paper, the possibility of new technologies which can be developed by integrating art and technology was discussed by referring to problems current technology is facing. It was also

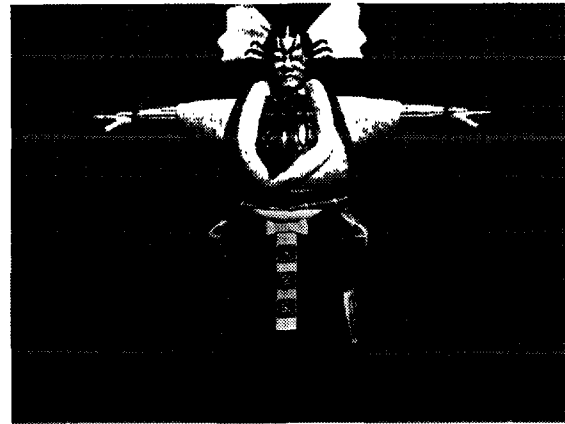


Fig. 7 An Example of a KABUKI Actor's Face Model

stated that, by combining AI research, such as processing technologies for image and speech, with artistic approaches, the possibility exists that fundamental technologies as well as systems and services for new communications can be created. As examples of this new direction, some projects have been introduced that are being done at ATR Media Integration & Communications Research Laboratories. These projects have just started but we highly await their results. Detailed progress reports will be given on other occasions.

## REFERENCES

- [1] N. Tosa and R. Nakatsu, "The Esthetics of Artificial Life," A-Life V Workshop, pp. 122-129 (1996.5) .
- [2] N. Tosa, "The Esthetics of Recreating Ourselves," SIGGRAPH'96 Course Note on Life-like, Believable Communication Agent (1996.8) .
- [3] N. Tosa, et al., "Neuro-Character," AAAI '94 Workshop, AI and A-Life and Entertainment (1994) .
- [4] K. Ebihara, et al., "Real-Time Facial Expression Detection and Reproduction System - Virtual KABUKI System -," Digital Bayou, SIGGRAPH'96 (1996.8) .
- [5] J. Ohya, et al., "Virtual Kabuki Theater : Towards the Realization of Human Metamorphosis System," Proc. of RO-MAN'96, pp. 416-421 (1996.11) .

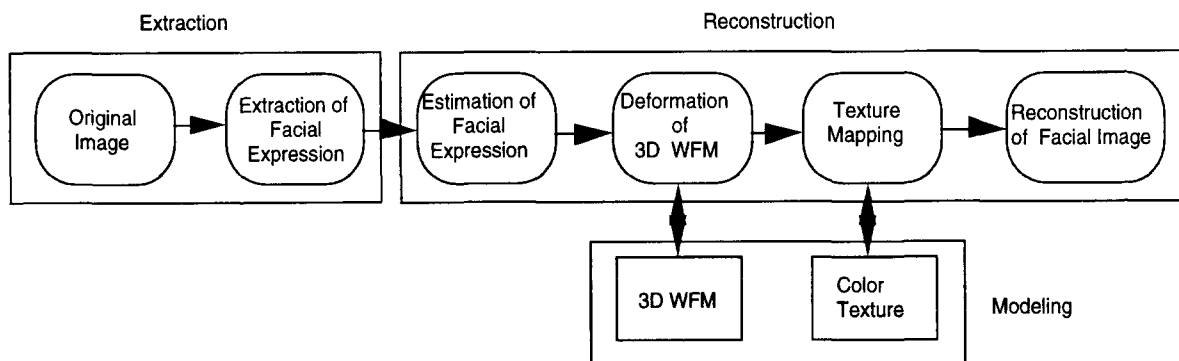


Fig.6 Flowchart for Reconstruction of Facial Image