Multimedia, Art, and Human-Computer Communications

Ryohei Nakatsu

ATR Media Integration and Communications Research Laboratories
2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288 Japan
nakatsu@atr.co.jp

Abstract

Although the conventional concept of communication is the exchange of logical information, the exchange of *Kansei* has become very important recently, especially in the case of mobile telephone communications. In addition, it has been recognized that the use of the body, such as facial expressions and gestures, is essential in communications. These facts indicate that technologies that can support *Kansei* expression/transmission will become key technologies for future communications. In this paper, several key issues related to *Kansei* communication technologies will be discussed first. Then, as examples of such technologies, research results obtained by the authors are described.

Keywords: multimedia, communications, nonverbal communications, art & technology

1. Introduction

Recently, the word *Kansei* has come to be used quite frequently, especially in the case of communications. The communications of today's young people via mobile telephones is sometimes called "*Kansei*-transmitting communications." In this case, they express *Kansei* by adding emotions and sensitivity to their voice. This is a typical method of *Kansei* expression. At the same time, the use of the body such as facial expressions and body motions is an important and essential means of expressing *Kansei* information in daily life. This is why face-to-face communications is the ultimate form of communications, and so far, no tele-communication means has been able to replace it.

These facts show that the means of expressing non-logical information such as *Kansei* is the basis of our communications and, therefore, technologies that support people in expressing and transmitting their *Kansei* will be essential factors for future communications.

For *Kansei* communications, there are several important issues to be considered. First, as we express our *Kansei* by using multiple modalities, the usage of multimedia is essential. Second, it is important to investigate why *Kansei* is essential for human communications. This will lead to the notion that even in the case of human computer communications the treatment of *Kansei* is indispensable. Third, as the treatment of *Kansei* is very difficult based on engineering approach alone, it is necessary for us to adopt new approach..

In this paper, first we will observe the trend of several media and it will be stated that in the future the integration of various media would occur. Also, investigation is made into the meaning that *Kansei* has in communications. It will be revealed that *Kansei* expression/transmission is the basis of our communications. Then as an approach to treat *Kansei*, a methodology called Art & Technology will be proposed. Based on the above discussion, two examples of research activities are described that are being conducted in ATR Media Integration & Communications Research Laboratories with the aim of developing technologies that support communications based on *Kansei* expression. Finally, an investigation will be made on the ultimate form of communicaions.

2. Trends of Communication Media

2.1 Multimedia and future communication me*dia*

What will be the future form of communications? We proposed "hyper-communications" as a new communications concept in the future(1). This was first based on the viewpoint that the boundary of communications media and other media would become ambiguous in the multimedia era and network era. Actually, such a movement is being caused by the various media (including communications). For example, a place for a new means of communications, the Internet, is being caused by the world of communications. It is thought that the Internet is a huge cyberspace that joins the whole world. People communicate with other people in that space and also shop.

In addition, looking at the movie industry, recent movies have been introducing digital techniques and computer graphics techniques and have been moving over to movies of a new generation. These

techniques have given the ability to create very realistic worlds, i.e., cyberspaces, in which expression had been difficult in conventional movies. Also, video games, especially role playing games (RPGs), have made it possible for people to enjoy a story by becoming the main characters in a cyberspace. From these trends, it can be predicted for "communications in a cyberspace" to become one keyword for the new communications (Fig. 1).

Fig. 1 New Trends in Various Kinds of Media.

2.2 Nonverbal aspect of communications

Next, our attention is to focus on the information exchanged in communications. Figure 2 shows a model of communications for human(2). In the surface layer, a layer exists that manages communications based on the use of languages. It is possible to say that research on communications and information processing has come to deal with the mechanisms in this layer. For example, objects that have come to be treated in speech recognition are the logic information included in sounds.

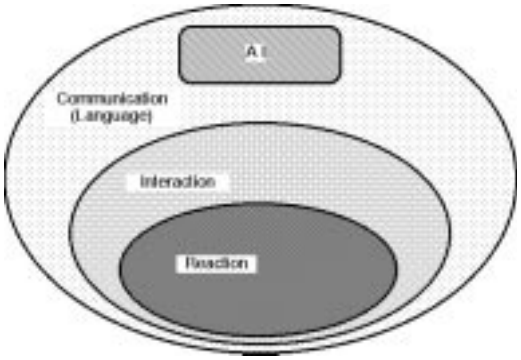

Fig. 2 Communication Model.

The logic information, however, is only a part of the information composing the sounds. Other rich information contained in the sounds include emotion information and "kansei" information. It is considered that such information is generated by a layer in a deeper level, i.e., the interaction layer in Fig. 2 or the reaction layer.

The interaction layer manages acts such as the changes in the speaking turns, for the rhythm and production of the utterances, to maintain the communication channels. This layer plays an important role for smoothly carrying out communications for humans. Below the interaction layer is the reaction layer. This layer manages the more basic operations of humans. For example, this layer has functions such as to turn the face to the location from which a sound had come or to close the eyes at the sudden appearance of light.

In this way, it can be considered that human communications plays a role with important functions of a deeper layer (in addition to logical acts and the treatment of information), and that non-logical types of information like emotions and *Kansei* are generated and solved through the movements of the above layers.

Conventional communications have come to apply logical communications, but in the future, other types of communications will become important. This, for instance, can be understood by observing the interest taken by young people in talking with portable telephones. In this case, it is possible to say that this phenomenon involves confirming sense-specific and emotions-specific relationships with partners and not the exchange of information with logical meanings. Consequently, the transmission of non-logical information will become another keyword of the new type of communications.

From the above considerations, communications to transmit non-logical information by using a cyberspace can be expected to be the form of the new means of communications. For our part, we are progressing with research on concepts and prototypes that will concretely achieve this.

2.3 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(2).

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.

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. 3.

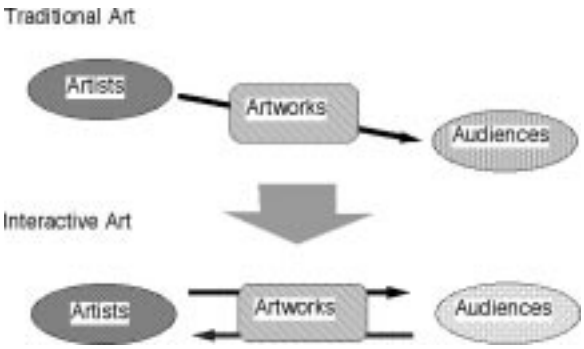

Fig. 3 From traditional art to interactive art.

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.

In the following chapters, two examples of research activities being conducted at ATR Media Integration & Communications Research Laboratories based on the previous considerations will be described.

3. MIDAS: MIC Interactive Dance System

3.1 Overview

Our research purpose is to establish a framework of non-verbal multimedia communication technologies by which we can express our emotional images directly. Art is an adequate reference for our research, because artists express their emotion by creating artwork. Dance is a form of art, and human motion is important in non-verbal communication. Therefore, it can be said that human motion analysis is important for our research purpose.

For this purpose we have developed an emotion extraction method for human body motion. This method can be applied to multimedia communication systems. MIDAS (MIC Interactive Dance System) is one of the applications. In this system, a performer can express his or her emotions by utilizing video and sound and can dance freely without any restrictions. MIDAS is unique as an emotional interaction system.

3.2 Overview of dance research

Choreographers have long been studying human motion in order to characterize it. A famous Austrian choreographer, Rudolf Laban (1879-1958), proposed three kinds of descriptions for human motion : Motif description, Effort-shape description and Structural description(3). The Motif description provides the most salient feature of a motion. The Effort-Shape description can describe human motion in terms of its quality and expression. The third kind is well known as Labanotation. Labanotation is used to systematically record human motion. It is useful in describing human motion in the technical research region(4).

The Effort-Shape description is adequate for categorizing human motion with emotional features. In this description, there are four parameters: space, weight, time and flow. Time has categories of sudden movement and sustained movement. Space stands for the design of posture and is the directivity of posture and movement. Weight stands for strength and the power of movement. Weight is categorized into firm movement and fine touch movement. Flow stands for the carefulness versus easiness that can be seen in movement.

In addition, another study has shown that dance motion patterns can be categorized into seven typical motives(5). The relationship between the Effort-Shape description and the seven motives is shown in Table 1. The seven motives represent the emotions of happy, solemn, lonely, natural, sharp, dynamic and flowing.

Table 1. Relationship between seven motives and Time-Space-Energy.

7 Motives	Time	Space	Energy
Happy	speedy	light	
Flowing	soothing		rounded
Lonely	slowly	sustained	
Natural	regular	sustained	balanced
Solemn	gradual	massive sustained	
Sharp	accented	firm sudden	straight
Dynamic	speedy accented	light	spread

The Effort-Shape description and the seven motives provide background knowledge for our research. We expect that Time-Space-Energy parameters can be obtained by image processing, and that the seven motives can be analyzed in subjective assessment experiments

3.3 Method of extracting emotion from human body motion

Our research purpose is to find a method by which we can determine the emotion contained in human motion. The concepts of Time-Space-Energy and the seven motives look to be useful for our purpose. The Time-Space-Energy parameters are somewhat emotional, but are close to the parameters that can be measured by image processing. The seven motives are useful for categorizing the emotions of human motion. Thus, we can directly estimate these emotions from video sequences if we relate the time-space-energy with the seven motives.

A professional dancer's scene captured by a camera was used for the emotion analysis. The dancer performed in accordance with the seven motives. Simple image processing was used to extract physical features from the dancer's sequences. To translate the physical features into emotional information, we applied a multiple regression method. Finally, we obtained a set of linear functions representing the relationship between the physical features and emotional information contained in human motion.

(1) Image Processing
As mentioned before, our purpose is not to extract accurate human motion parameters at joints. Our purpose is to abstract human motion. Thus, human motion should be measured with a simple method.

We use the center of gravity and the circumcised rectangles of a dancer's silhouette images as shown in Figure 4. The center of gravity represents the dancer's motion over the dancing space. The circumcised rectangle represents the human local motion(6).

Fig. 4 Silhouette images of reference dancer.

(2) Time-Space-Energy
We re-define the Time-Space-Energy features as follows.
Time: Human motion speed
Space: Openness of human body
Energy: Acceleration of human motion
We assume that we can represent Time, Space and Energy with a linear combination of physical parameters obtained by image processing. The Time-Space-Energy concept originally came from research into choreography. Therefore, the Time, Space and Energy functions should be examined according to psychological factors. Subjective assessment tests are conducted to examine the psychological characteristics of each sequence. Based on a multi-regression analysis using the psychological test results and physical parameters, the linear combination relation is determined.
(3) Seven motives
We need to clarify the relationship between Time-Space-Energy and the seven motives. Other psychological experiments are conducted to categorize the reference dancer sequences. In these experiments, subjects vote for one motive out of the seven motives for each reference dancer sequence. We assume that the seven motives can be represented by linear combinations of Time, Space and Energy. The multiple regression method is used to determine the linear combination.

With these functions, we can determine one motive among the seven motives. The highest value is chosen to select one motive among the seven

motives.

3.4 MIDAS

(1) Multimedia system and its content
The extracted emotional information is sent to a multimedia controller. The multimedia controller manages a video switcher, real-time disc system, and sound system. The multimedia controller interprets the received emotional information and selects adequate video and sound clips. The video clips are displayed on a 120-inch projection monitor. The sound clips are played simultaneously. Figure 5 shows a schematic of MIDAS.

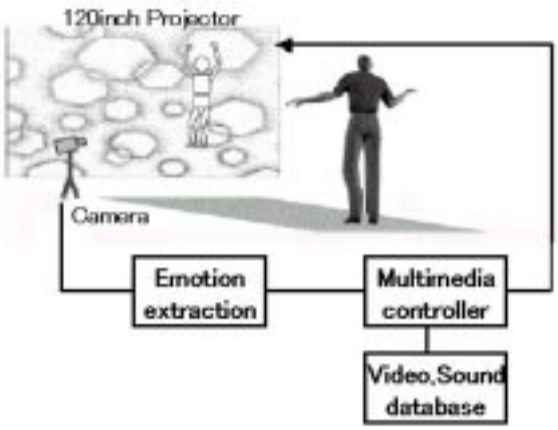

Fig. 5 MIDAS schematic.

A performer's image is synthesized on video clips. Time-Space-Energy parameters are used to make the performer's image. Figure 6 shows the appearance of MIDAS.

(2) Consideration of MIDAS
The performer can enjoy the dancing and interacting with the video and sound. As previously mentioned, our research purpose is to express emotional images by using a multimedia application. MIDAS is a multimedia system for expressing emotion with human motion.

Fig. 6 Appearance of MIDAS.

A dancer usually dances along with music. When a dancer dances in MIDAS, he or she can change video clips and hear sound clips. This means that the dancer can lead the music and the video, so he or she may feel strange. However, according to interviews with users, they have fun with MIDAS. We believe that there are some reasons for this fun. The first reason is that they are requested to move their bodies for interaction. This gives them a more immersive feeling. The second one is that they can see themselves on the monitor in front of them. This may also have immersive effects. The third one is that there is no constraint on their movement, that is, they do not need to pay attention to markers or sensors.

3.5 Future work

MIDAS has many aspects in addition to a dancing system. Some people say that this system can be used for psychotherapy. Educational applications can be considered too. Of course, MIDAS can be changed to an interactive multimedia game by replacing the video and sound content. We are planning to develop an application for amusement. Figure 7 shows an example of its content, where a performer can control the avatar motion.

Fig. 7 MIDAS for amusement applications

This means that several performers will be able to use MIDAS. In the current MIDAS, only one performer can use it, because it may be impossible to represent the emotions of plural performers by using a set of video and sound. In addition, we can extend MIDAS to a telecommunications system where performers can dance over a network.

4. COSTUNE

4.1 Overview

Here, we describe a wearable musical instrument called "CosTune" (costume + tune) that is equipped with wireless communications functions. We discuss how CosTunes can help to enhance communications and to form communities in the real

world.

Music plays an essential role as a communications medium. For example, the members of a jazz band communicate with each other by performing music, and the band conveys a certain impression to the audience by their music. Another example is when we travel to a foreign country; we can communicate with local inhabitants by singing with them at a public house, dancing to music at a festival and so on, even if we cannot understand the language of the country. Thus, music is essentially a powerful communications medium as well as a universal language.

Several wearable musical instruments have been developed, e.g., YAMAHA MIBURI™ and BODYCODER(7), although they are still restricted to use on a stage. The Musical Jacket(8) developed at the Media Lab., MIT, achieves true portability. However, the authors see it as a simple extension of ordinary musical instruments; sound is still the only vehicle for communications and the aspect of a communications tool is not apparent in its design.

We designed "CosTune" to be a communications tool rather than a simple musical instrument. The most significant feature of CosTunes is that they are equipped with wireless network functions. This allows users to communicate with anyone, anytime, and anywhere by means of music in an ad hoc manner without scattering sound. We think that this feature makes CosTune a supporting tool for forming communities in the real world.

4.2 Conceptual system design

The system consists of CosTunes and servers. A CosTune is a portable musical instrument that can be worn by a user. A server is located somewhere in town and communicates with CosTunes that are passing by the server.

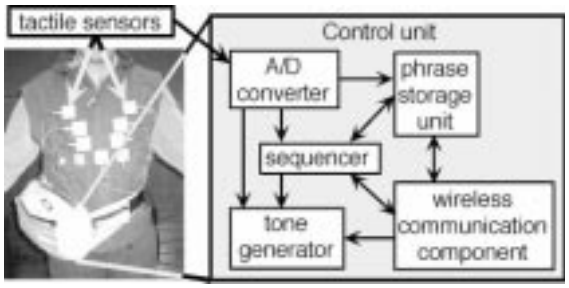

Fig. 8 Componets of the Costune

(1) The CosTune
The CosTune consists of a wearable input device and a portable control unit (Figure 8). The wearable input device is a cloth on which a number of tactile sensors that correspond to, for instance, the keys of a piano are mounted. By manipulating the sensors, the user can play music. The portable control unit is equipped with an A/D converter, a tone generator, a wireless communications component, a phrase storage unit and a sequencer.

The output signals from the sensors on the wearable input device are input to the A/D converter that converts input analog signals into MIDI data. Currently, the mapping between a sensor and output MIDI data is stationary. However, variable mapping based on, for instance, the "fixed note-function mapping concept" (9) is also useful to support novice players, in particular. The output MIDI data from the A/D converter and the sequencer that plays back stored phrases in the phrase storage unit are input to the tone generator. The output sound signals from the tone generator are input to the headphones. In the meanwhile, if the user wishes, the performed phrases, i.e., the output data from the A/D converter, are stored in the phrase storage unit.

The wireless communications component transmits as well as receives "phrase packets" to and from other CosTunes and servers. The phrase data, as MIDI data, that are obtained from the received phrase packets are also input to the tone generator as well as, if the user wishes, being stored in the phrase storage unit.

(2) The servers
The servers are located in places around town like the base stations of a cellular phone. The constituents of a server are almost the same as the portable control unit of the CosTune, but the A/D converter and the tone generator are not necessary. The role of the server is to store sets of musical phrase data and to exchange them with the CosTunes. Therefore, the phrase storage unit of the server should be larger and more intelligent than that of the CosTune.

(3) Phrase packets
The CosTunes and the servers exchange phrase packets. One phrase packet includes the following items:

- Phrase data,
- Length of the phrase data,
- Attributes of the phrase data, e.g., rhythm, tempo, timbre, a musical genre, a part in the musical structure,
- An owner's profile, e.g., owner ID, age, sex, musical preferences and place of residence,

where, a phrase is a certain length of song component, which can be divided in time and/or instrumental type or role. Conversely, a song is defined as an

organized set of phrases. Phrase data are sequences of symbols representing the notes that the phrase consists of. The MIDI format is usually employed to describe phrase data.

The relations among the musical genre, the part in the musical structure, and the timbre should be defined beforehand and shared by all CosTunes and servers. So, when a CosTune is going to exchange a phrase with another CosTune or with a server, one adequate phrase is chosen according to the relation. If a user indicates one specific genre and one specific part, his/her CosTune can collect all parts of the genre except for the indicated part, assemble them simultaneously, and form the accompanying data of a song for him/her. As a result, so-called "minus-one" data or "karaoke" data can be adaptively obtained.

(4) Requirements of ad hoc networking

The ad hoc networking function is very important for the CosTune system. Figure 9 illustrates an example of an ad hoc network where two network segments already exist, i.e., segment A with four CosTunes and segment B with three CosTunes and a server.

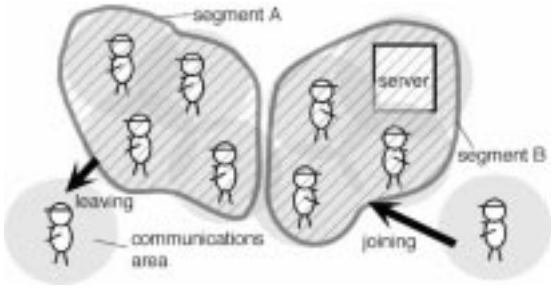

Fig. 9 An example of ad hoc networking

Within a certain segment, all CosTunes and servers must be able to bi-directionally communicate with each other and multicast phrase packets through a shared unique channel. A CosTune user must always be free to join or leave a segment at any time. For example, in Figure 9, a CosTune user has just left segment A while another CosTune user is about to join segment B. In addition, CosTune should not be involved in multiple segments at the same time. This is because a user usually finds it hard to play and/or to listen to multiple performances simultaneously. Therefore, a segment should be exclusive.

The communication areas of the CosTune and server should be comparatively small. A 20 to 30-m-radius area is suitable for a CosTune while a 50 to 100-m-radius area is suitable for a server. Larger CosTune areas, in particular, make it very difficult to find who the co-players are when the CosTune user has a jam session. Since the users may dance while performing, occlusion-free and non-directional communication is necessary.

4.3 Applications and protocols

With only one CosTune, a user can individually play music by herself/himself while walking on the street or dancing. However, using the ad hoc network function, new musical activities such as an ad hoc session or ad hoc collaborative composition can be achieved.

The ad hoc session, for example, enables real-time, live musical entertainment, which is one of the main pleasures of playing music. Someone who wants to play in a session looks for partners by strolling around town while using this function. When the wireless communications areas of some CosTunes overlap, if the users' musical preferences are similar, and if all of them want to have a session, an ad hoc real-time musical session can be immediately started there even if they are meeting for the first time. Needless to say, an existing band (where the members of the band already know each other) can form a street performance with this function, too.

4.4 Prototype system

In order to preliminarily evaluate whether we can actually perform songs while walking and whether we can form a session with a number of players using a wireless network, we created and tested a prototype system. The prototype CosTune is equipped with all of the components described except for the sequencer (Figure 10). We used the i-cube™ system, which is an infusion systems™ product, as the A/D converter. The i-cube system converts input analog signals into MIDI data.

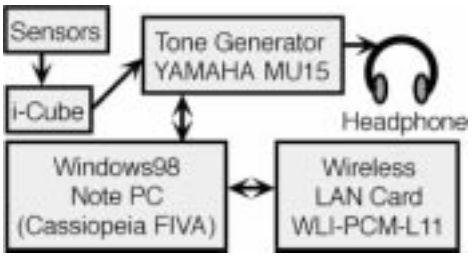

Fig. 10 Construction of prototype CosTune.

On the contrary, the prototype server is equipped with a sequencer. We prepared three types of interfaces: a jacket type, a pants type and a glove type (Figure 11). Although any kind of sound can

(c) Glove type CosTune

be assigned to any type of interface, we assigned an organ sound, drum sound and strings sound to the jacket, the pants and the glove interfaces, respectively. As the wireless communications component, we used wireless LAN cards (IEEE802.11b conformance, 11 Mbps). Currently, simply encapsulated MIDI data are transmitted as UDP packets among the CosTunes and the server. Neither the "phrase packet" exchange nor the protocols for the ad hoc collaborative composition and ad hoc session have been implemented yet. The transmitting delay is less than 10 ms. This is short enough for most of the amateur players to play without confusion.

(a) Jacket type CosTune

(b) Pants type CosTune

Fig. 11 Prototype Costune.

On each CosTune, performance data are converted into MIDI data and are transmitted to the server encapsulated in UDP packets. Then the server corresponds the received data to a specific timbre based on the MIDI channel (a unique MIDI channel is assigned to each CosTune) and immediately broadcasts the received data as well as accompaniment data that are generated by the sequencer of the server to all of the CosTunes. Every CosTune receives the broadcasted data and inputs them to the tone generator. Thus, every player can listen to the performances of all of the players as well as the accompaniment through headphones.

4.5 Discussion: CosTune as communityware in the real world

There are several systems and projects that allow people to perform remote sessions by exchanging performance data in MIDI format over the Internet (e.g., TransMIDI[7]). CosTune is similar to these attempts in terms of phrase exchange through a network. However, as for musical performances, we regard communications in the real world as the most important while communications using systems are done in virtual worlds. Therefore, we insisted on a wireless communications component with a comparatively narrow communications area. Additionally, we insisted on a wireless ad hoc networking function that always automatically looks for other users who have similar musical tastes. These allow people to meet in the real world and to have jam sessions in a face-to-face manner; an essential joy of musical performances.

Moreover, we are interested in the characteristics of the "areas" of cities, towns and so on. Specific kinds of people tend to gather in specific kinds of areas. They generate the "atmosphere of the area" and the atmosphere of the area attracts those who like it. As a result, areas acquire unique characteristics, e.g., SoHo in N.Y. and Harajuku in Tokyo. We think that the music that is performed in an area must reflect the characteristics of the area. Conversely, the jam sessions and composed musical pieces must become different depending on the areas where a CosTune user visits. Therefore, we think people who want to enjoy the music of a specific area should actually visit the area and meet the people of the area.

5. Ideal Communications

In the previous chapters two examples of new communication systems that would activate our *Kansei* expression and transmission were described. To progress with research on such a new means of communications, it is important to consider how the

ideal type of communications should be. In the conventional concept of communications, logical information is sent from a sending side to a receiver. In the following phase, the sender and the receiver alternate their roles, and information is sent in the opposite direction. It cannot help but be said that such shear non-style communications is indeed the exchange of static information.

It can be considered that the original form of communications was more dynamic. The sender and the receiver, and the environment surrounding them become united, and it is possible to understand the information as something shared between the sender and receiver, and moreover, as the relationship of the sender and receiver, and not as something handled as a physical amount separated from them.

From such a viewpoint, it can be thought that the elements the ideal communications should have include the following three items.
(4) Experience and synesthesia
(2) The integration of experience involving the body and mental experience
(3) The ability of active immersion

Below, explanations are given for each of the three items.

5.1 Experience and synesthesia

Communications to date have meant the sharing of information (related to communications) between people, by having logical information be sent and received. This, however, has indeed been "dry" communications.

The communications between a mother and her child, for instance, is close communications - both see and hear the same things in the same environment, and moreover, share the same feelings. Concerning this, more than to say the sharing of information, is it possible for one to call this a condition whereby a partner's feelings and one's own feelings are synchronized by having both experience the same experiences? Consequently, we want to point out "experience and synesthesia" as being the state of the original communications.

Concerning the term "experience", in addition to the meaning of "obtaining and sharing the same information", is not the implication strong for it to mean "to get information by moving the body"? Then, it should be possible to consider "experience = information + body". Next, concerning the term "synesthesia", it is possible to consider it as "strong thoughts about one's partner who has the same ideas and experiences". At this time, it can be considered that the "resonance phenomenon of some seeds" occurs between both sides. Consequently, it can be said that "synesthesia = sharing + resonance".

In other words, it is possible to say that the condition under which mutual thoughts are sympathized by the sharing of the same experiences, i.e., "experience and synesthesia", involves shallow communications; this is against the thinking of the conventional shallow communications (that is, to merely share logical information).

5.2 The integration of experience involving the body and mental experience

Next, we consider the contents of "experience". "Experience" is created from experience involving the body and mental experience (Fig. 16). The former merely means to move the body. A representative example is exercise, but even a thing moving only a part of the body is experience involving the body. For example, experience involving the body includes the emitting of sounds. Moreover, even the action of viewing is an experience involving the body when attempting to view an external world because of the accompaniment of operations to move the head to the desired viewing direction, and so on.

In the past, only those operations involving aggressive body movements were called experiences having to do with the body, but here, we want to call actions having relationships to an external world by the use of the five senses including the actions of viewing and listening (which have come to be considered passive), as experiences involving the body.

For the latter (i.e., mental experience), it can be considered that the functions of the cerebrum become involved in actions added to the functions of the five senses such as using words, reading books, listening to music, and looking at pictures. Of course, here the handling of logical information occupies a large space, but the handling of "kansei" information such as emotions and feelings also occupies a large space.

It is in mental experiences when treating "kansei" information that we have an interest here. It is possible to say that all of the following are mental experiences of a high degree: expressing ones feelings by using words, becoming moved by reading a book, being impressed by listening to music, and becoming immersed in the world of a movie while watching it. The original meaning of the word "experience" involves an "integrated experience" merging the above experience involving the body and

mental experience.

The act of creating art like drama, performance, and sculpture has a side including experience involving the body - in that it achieves form by a practitioner being able to appreciate a high mental experience and moving the body with it. From this, it can be said that one is able to appreciate an "integrated experience" (combining experience involving the body and mental experience) when involved in the activity of the creation of such art. As a result of such an integrated experience, a kind of fascinating situation to appreciate is obtained during the art creation. Considering the above in this way, it becomes possible to explain why the media "karaoke" has become a popular media.

First, "karaoke" is an experience involving the body from the point of producing sounds. Moreover, it is also a mental experience from the point of reading and understanding words, i.e., the lyrics of songs, appealing to the emotions and feelings of humans. Consequently, it is possible to say that "karaoke" is a media to make one appreciate an integrated experience. Is there not a sense of fascination felt when singing "karaoke" that connects with and fits the sense of fascination an artist feels (said to be of a low level)?

In addition, as is generally said in many conversations, the mental aspect becomes more and more important for a sport when one enters the professional domain. This applies to baseball, among other sports. Therefore, if a sport reaches the professional domain for one by crossing over from the level of "fun", it is possible for that person to reach an integrated experience (i.e., experience involving the body and mental experience).

It is said that sports players often feel a sense of uplifting and a sense of fascination, but are these not the same as the sense of fascination felt when singing "karaoke" or the sense of uplifting an artist feels when creating art work? As explained above, an integrated experience is an experience called a "kansei" of a high dimension, and it is desirable to be able to achieve such an experience in communications among human beings.

5.3 The ability of active immersion

In an integrated experience, things able to be felt raise the sense of uplifting and the sense of fascination, but will anything be okay if they take the trouble of showing these? From this point of view, the character "Pocket Monster", which has recently become a topic of conversation, has become a kind of fascination and uplifting. There probably are various kinds of fascination and uplifting, but here, the word "immersion" is used as a general term for these. When looking at the Pocket Monster incident (i.e., many children being affected by seizures), it can be understood that there are many types of immersion. Here, we consider them by classification into "passive immersion" and "active immersion".

Does not the level of consciousness differ between both, although there is no change in becoming absorbed in both passive immersion and active immersion? Is there not a big difference for the condition of fascination, that is, the condition of losing consciousness by being mesmerized by viewing Pocket Monster? Consequently, the condition where one completely forgets oneself or the condition where one loses consciousness (fascination, hypnosis, trance) becomes a key word that explains the feeling of passive immersion.

Active immersion, in contrast, makes possible the condition whereby one's consciousness is maintained in a normal manner when that person becomes absorbed. Conditions of immersion while clearly maintaining consciousness include the condition of doing work with concentration and the condition of immersion in the activity of creating art. Even in the case of sports, such conditions are experienced.

Do not these conditions involve expressing the condition of immersion for a professional baseball player or a professional golfer participating in a game, assuming "the nerves are able to concentrate on only the ball in front", "the ball thrown by the pitcher seems to have stopped", or so on? In this case, the consciousness is extremely clear and the actions of the person are able to be controlled completely. Unnecessary idle thoughts, however, are excluded.

When such passive immersion or active immersion pays attention to the processes that take place, an interesting fact becomes clear. It is the existence or non-existence of interaction.

Passive immersion does not work on a partner only by the receiving of information, like with Pocket Monster and mesmerism. In other words, there is a lack of interaction here. In contrast, active immersion differs from the point of working on the object, like with art creation and with sports. In other words, interaction exists with active immersion. Consequently, it is possible to say that the existence or non-existence of interaction is the key that distinguishes between passive immersion and active immersion.

Therefore, if interaction exists, will this result in

active immersion? Unfortunately, the situation is not that simple. For example, let us consider the case of interactive art. Interactive art is said to be art where the picture or sculpture itself changes by movements, such as by the body and hand movements of the audience and their voices.

Novelties that interactive art asserts in comparison with conventional art are the achievement of active immersion by the intervention of interaction and the possibility of a higher appreciation for art than before. However, there is also a lot of interactive art that pulls the interest of the audience at the start but soon loses the interest of the audience, without leading to the achievement of active immersion. Is not a large reason for this because interest leans only towards the introduction of interaction and there is no deep insight into whether or not one aims at achieving something through the introduction of such interaction?

6. Conclusion

In this paper, we described communication for exchanging *Kansei* information and introduced several systems that support people in expressing their *Kansei*. First we investigated communication and pointed out that *Kansei* communication will become essential for future communications and telecommunications. As a result, we made it clear that the technologies that support *Kansei* expression are key factors for future communications. Also we proposed a new approach, called Art & Technology, to develop technologies *for Kansei* communications.

Then we introduced two systems that are being studied by the authors and that are good examples of such technologies. First we described an interactive dance system called MIDAS. In this system, the dance movement of a user is analyzed and classified into one of seven typical dance motives. According to the result the system can generate music and images that fit the motion of the user. Then we described the design and applications of CosTune, a music-mediated communications tool. CosTune is equipped with several sensors mounted on clothes as well as ad hoc networking functions to exchange phrase data and user profile information with other CosTunes and servers that are located around town. CosTune will allow a user to hold an ad hoc session on a street corner with other users who share similar musical tastes even if they just happen to meet there and do not know each other.

Finally as an ultimate target of our research, we have investigated the ideal form of communications. We have pointed out three key factors for the realization of ideal communications and investigated in detail each of these factors.

Although the study of technologies that support *Kansei* expression is in its initial stage, the prototypes we have developed show good performance and we believe that this area will become one of the central areas for the communications research.

References

(1) Nakatsu, R. (1998). Toward the Creation of a New Medium for the Multimedia Era, *Proceedings of the IEEE*, Vol. 86, No. 5, pp.825-836.

(2)Nakatsu, R. (1997). Image/Speech Processing that Adopts an Artistic Approach -Integrating Art and Technology-, *Proceedings of ICASSP1997*, pp.207-210.

(3) Hutchinson, A. (1996) Labanotation. In: *Dance Books*.

(4) Nakata T., Sato, T. and Mori T. (1998). Expression of Emotion and Intention by Robot Body Movement, *IAS-5, IOS Press*, pp. 352-359.

(5) Matsumoto, C. (1987). Dance Research : Problem Situation and Learning of Problem Solving II – Qualities of Movements and Feeling Values, *Proc. Japanese Association of Physical Education and Sports for Girls and Women*, pp. 53-89, 1987 (in Japanese)

(6) Iwadate, Y. (1999). Study on Image Expression, *IEICE HCS99-53*, pp.87-94 (in Japanese).

(7) Bromwich, M. A. and Wilson, J. A. (1998). BODYCODER: A Sensor Suit and Vocal Performance Mechanism for Real-time Performance, *Proc. International Computer Music Conference 1998*, pp. 292-295.

(8) Fels, S., Nishimoto, K. and Mase, K. (1998). MusiKalScope: A Graphical Musical Instrument, *IEEE Multimedia Magazine*, July-September, pp.26-35.