

**International Congress of Aesthetics 2007
“Aesthetics Bridging Cultures**

**Double Visualization and Multiple
Interactivity in Music ---towards the
situation of Japanese live-electronic
audio-visual musical piece**

*Mikako MIZUNO , associate professor of
Nagoya City University, Japan*

This paper tries to present one phase for analysis of musical interactive pieces, especially audio-visual musical piece. Some models for on-stage interactivity are discussed, and classification is presented based on some Japanese examples including DIPS, jitter etc. Here, on-stage interaction means the realtime audio-visual interaction during the musical performance. In that situation composers can make new audio-visual interactive contexts, which are different from those of traditional movie or film. In this paper, the new type of interactivity is called <multiple> and the situation that includes such kind of interactivity is described as <double visualization>, because the audience can perceive both the performer's musical action which should generate sound and the image as a result of computing. In the paper, the interactivity is not classified as the binominal chart of sound and image.

Introduction : What is Interaction in the Audio-visual Pieces?

Since the latter half of 1980's when the information technology proceeds to operate on the realtime sound data processing on the personal computer, the term <interactive> has been an important keyword for the computer music in the point both of technology and of aesthetics. It is the co-operative research of music and science in IRCAM in Paris that has become the origin of the on-stage live electronics with realtime sound processing. And today's Japanese situation of that kind of musical pieces are strongly affected by the system developed in IRCAM until 1984. The system kernel is the interaction described with the language MAX.

Interaction means two-way communication between two different class of things or persons and interaction in musical pieces has been discussed mainly as relationships between musical gestures and sound generated by the performer's gestures as in some branches of computer science. In other words, in such scientific branches, researches of interactivity of musical pieces have been focusing on action/sound relation as information processing causality. Some instrument-like interface products have been also discussed as tool which mediates gesture and sound.

Here in this presentation, today's situation of computer music pieces performed on-stage is discussed from the view point of interactivity; what kind of musical contents is to be transmitted from which factor to which. This concerns both about the information data flow and about communication aesthetics of music.

I will discuss some Japanese interactive music pieces which were mostly effected on IRCAM-originated realtime sound processing, especially the technique using music descriptive language MAX. I do not take technical concepts like interface design, input device and mapping which was reported in the discussions by Marcelo Wanderley or some conference papers like in New Interface for Musical Expression(NIME) or Sound and Music Computing(SMC). The interaction between the physical quantitative parameters or information technologies like sound, movie or gesture will not be discussed as concepts, for these are the file code types which are dominant parameter units as information file types. They are indispensable and the important materials in the process of composing in order to construct the structure of the piece, but they are hidden structure for the audience as the twelve-tone row in the dodecaphony. Discussion about interactivity in the audio-visual musical pieces on-stage should examine not only gesture-sound, gesture image and sound-image relationships but also composition-performance and creation-reception relationships. The general data flow and communication situation is illustrated in [Figure1] .

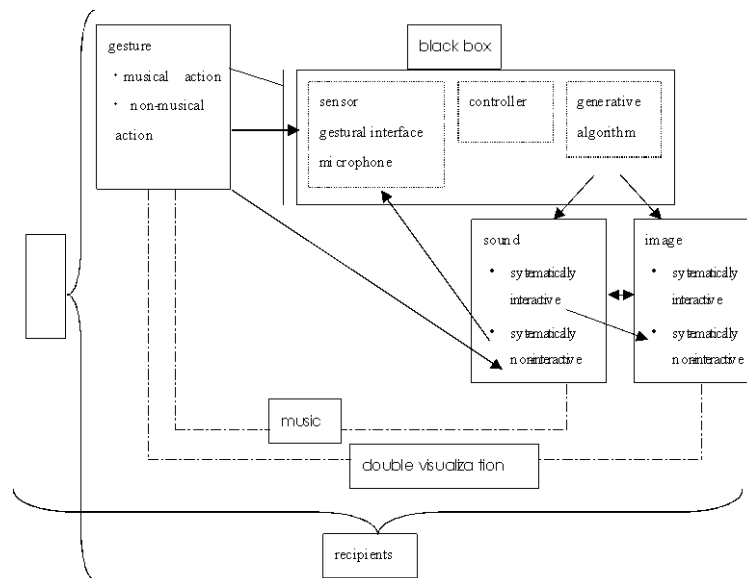


Figure 1. General Situation of Interactive Computer Music

Here is a multi-phase and somehow nesting structure representing the situation of audio-visual interactive music.

The first phase is the technological interaction as the digital data flow. This technological flow is based on the framework of the sound control parameters like sound files of <.wav>, <.aif>. This is a kind of bundle of signals which decide the profile of the sound, and does not necessarily decide the character of the music. This kind of data flow is to be <encode> or <decode> for the purpose of processing. The composer knows the function of the processing, but the recipients do not know.

The second phase is the fundamental communication, which is based on the traditional human-to human relation in music ensemble. Ensemble interaction is imitated as simulation in today's computer music. For example, the <score-following> patch in Max is the simulation of the musical situation where a conductor is leading all the performers

on the stage. But most of the ensemble interactivity cannot be digitized directly because the complex phases of music ensemble are not so much researched because of the ambiguity of the articulation of the structural elements. Only some phases like tempo unity, tempo change, starting point of the phrase through some physical actions by the performers, dynamic change can decide the expression of the performance, which make signs from one performer to the other. They are the performance phase as musical contents. The informational parameter units of music ensemble can be highly digitized by way of each composer's musical thought about music ensemble, performance and the music itself as one art genre. This is the composition for performance not only in musical score but also through system and environment design for the new audio-visual communication situation. The essential ensemble interactivity will be discussed later.

As for communication, we should also think the situation of the audience.

The audience, as recipient of the created piece, see the performers on the stage, the instruments, the visual images on the screen and hear the sound generated from the instruments(voice) and the loudspeakers. The recipients sometimes conceive the relationship between the performer's actions and the sounds and images, because they compare the on-stage situation with the ordinary type of music performance.

I dare to say the word<comparing> to discuss the individuality as Jean=Jack Nattiez used the word <seriation> (mise en serie) to analyze individuality compared to another pieces keeping the similar style.

Even though, in the case of in-exhibition pieces like installations, which are not the objects of my present paper, the recipients compare the corresponding situation of input and output with that of ordinary environment.

The recipients see and hear the gesture of the performer(s), sounds visual images like graphics/video and their reciprocal relationships, i.e. the causality between the events and the sounds/images as the results of those events. The recipients conceive the relationship between the performers' actions and the sounds and images. The system is hidden from the stage but the recipients know some data flow by experiencing the on-stage audio-visual environment.

Visualization as for Music

Now I discuss the multiplicity comparing with the fundamental situation of musical communication. Here I call it <double visualization>. The visualization is sound-generating on-stage audio-visual communication. The word, double, signifies there are two different levels of audio-visual corresponding relations. One is that of the traditional ensemble with which the audience can know how the sound generated by the performer who are communicating with each other or make their musical thinking mediated through the instruments or their bodies. Ensemble can be both among the performers , and, between the performer and the sounding body i.e., instruments.

The other visualization is non-sound-generating, that is, the digital images on the screen. The screen is on-stage and face to the audience but it is a different space from that of on-stage live performance , space which is shared with audience in realtime. In the <gesture box> in [Figure1], there are two items: musical/non-musical gestures. Both of the two items have two types; sound-generating and non-sound-generating.

So, if we set parts of music gestures as actions, we can make 4 types of interaction between actions and sound.

- 1... sound-generating musical action
- 2... non-sound-generating but musical action
- 3... sound-generating but not musical action
- 4... non-sound-generating and neither musical action

The first type includes the traditional ensemble performance on-stage, and the sound generated both by the performers and by computer. In the case of sound-generating, it is difficult for the recipients to distinguish the sound which is coming by way of the <black box> from the sound which comes directly from the performer, especially in case that the performer has a musical instrument and the <black box> is generating similar sound to the instruments, for the audience can guess the origin of the sound only through the audio-visual causality. Here the causality is blurred.

The second and the third types are the environments where musical actions do not any connection with the sounds. In the third type, not the musical actions but another actions, for example, telephoning action, switching the TV channels or generating the sounds on-stage. The sounds generated by these non-musical actions are not the direct sounds resulting from the actions but exchanged to another sounds through the black box of the program, and the sounds, otherwise only the noises in the ordinary life, become the sound elements composing the music piece. One example of the second type is the action of playing the violin that is to be exchanged to the conversation in the ordinary life, which make the pizza shop deliver the pizza. This is the case of the Formant Brothers¹, Nobuyasu Sakonda and Masahiro Miwa.

One example of the fourth type is the piece "Timetracks" (2005) by Masayuki Akamatsu², in which video images are printed on the paper by mapping program. There are no musical actions, sounds either.

Are these interactive pieces music? Music is not the sound itself. If the logic dominating the piece comes from musical ideas, the piece may be musical piece even though there is neither sound nor musical actions. In the pieces of Akamatsu and Formant Brothers the logic led from musical ideas are basically communication in real life as conversation. And the conversation comes from ensemble in music. It is necessary here to analyze the ensemble as conversation depending in music. Akamatsu himself says <Timetracks> as media art, and not as music. We have discussed on-stage musical performance including the human-to-human, or human-to instrument ensemble. So here we will rethink the traditional ensemble situation.

6 categories of the interaction features in the traditional musical practice

I focus on six categories that cover the communication situation of the traditional musical ensemble and which distinguish two phases of performance; one is thinking or responding on/to given information during the stage performance and the other is making or generating sounds.

- 1 . responding after listening: given information is audio information.
- 2 . responding depending on the precedently determined information like meter, regular tempo and so on.

- 3 . responding depending on visual information like body motion of the other players , or some non auditory information like breathing, judging on the knowledge of accumulated information about the current performance, for example one player feels the other player's tendency of acceleration. The response and the action of generating sounds are occurring in real-time along with the time sequence of the performance, but depending on non auditory information, whether they depend on some precedently determined information or not.
- 4 . responding as making/generating sounds.
- 5 . responding as recognition of receiving information: reception of the information is to be presupposition for sound generating onward. The recipient of the information (performer) is not going to make action for sound generating. He/She receives information just as necessary knowledge, like meter and tempo of the current performance.
- 6 . responding as preparation for immediate sounding; not making sounds immediately but making preparing action for keen sound-making.

Theoretically interaction in musical ensemble may have six features corresponding to these six categories. Technically speaking interaction in music performance should be tested in two phases of information processing, e.g. receiving and sending information. The 1,2,3 categories above mentioned concern the types of information which are received and conceived by one player who is in the midst of music ensemble. They are also three classes that make up the collection of sound-generating responses. The given data in the second categories are non-auditory information. The 4,5,6 categories concern the types of responding actions which come out from the sound-generating performer. The given data in the fifth and sixth categories are non-auditory information.

According to this classification of traditional interaction phases, the possibility of audio-visual interaction with today's digital processing in musical pieces can be set in more sophisticated framework. The recipients of the piece see and hear the situation where the performers receive and make out auditory and non-auditory, visible actions in the traditional on-stage ensemble. This is also visualized situation on-stage conversational communication. Therefore, the recipients experience the audio-visual space which includes the traditional ensemble conversation and systematically made, new audio-visual causality space.

The recipients of the piece can experience the pieces mainly through the auditory perception. And the auditory perception in the concert, where the musical performance occurs on the stage of concert hall or some performing places, accompanies visual perception which affects on cognition concerning about the causality between visual and auditory information. The recipients of the audio-visual musical pieces on-stage see the gesture and some other musical events in hearing the sound and at the same time in seeing the image on the screen. To hear is connected to the area of the action-sound causality and to see is connected to the area of the action-image causality. These causalities are technically realized through interactive system like sensor/interface controlling generative algorithm, and the recipients can experience the interaction with or without understanding of the system construction.

As the interactive system for itself is concealed from the recipients, the system is a black-box for them. They can perceive the system interactivity only when they can see the

gesture of the performers and see/hear the image/sound with reference to their ordinary experience in the ordinary space; they presume what kind of action leads what kind of sound. In the case that the recipients can perceive the causality from gesture to sound-image on the stage, the interaction can be said as <visible interaction>.

Actually the images as a result of interactive system are projected on the screen on the stage. So the recipients see the images as play within a play. How relates the play within the play to the primary play of the performer in the aesthetic point?

If we call the visible interaction between the gesture and the sound <interaction 1>, the visible interaction between the gesture and the image <interaction 2>, and the invisible technical interaction between sound and image <interaction 3>, we should present the fourth one as <interaction 4> between the total sound and the visual construction of the total stage including both the screen and the performing actions. The total audio-visual stage includes double visualization, even if the recipients do not know the multiplied communication.

Now we reach to find another phase of the problem, if we set the trilogy of the gesture, the sound and the visual construction on the trilogy of the ontological communication chain between composition, performance and reception. How do the composers arrange the four interaction in his creative process? And how is the multiple interactivity conceived by the recipients? How the performer work as media of the interactivity? It is these problems that are to be answered by the composers of the interactive music.

Composition process comprise both the instructions to the performers with or without score and the system design which sometimes comprises algorithm generating sound and music structure. Composer's instructions with or without structural algorithm are to be presented to the performers. Performers as recipients of the composer's instructions have some intentions about their performances. Their intentions are affecting on their performing actions in various levels, for the performing actions are consisted of gestures to make sounds, preparatory motions for the following musical events (phrase or making sound) and communication response for the musical ensemble. All or some elements of these performing actions are mediated by computer and transmitted to sound/image processing.

Gesture of Musical Action

Gestures controlling sound or image through some interactive program can be classified into some communication stages on the points of musical meaning; preparatory motions for the following musical events, ensemble communication or communication response for the musical ensemble and sound making action. Most of the cases of today's interactive systems capture the performer's different actions as same level information without regard to to which stage of the musical meaning they belong. The more similar is the way of the reflecting system of the black box from action to sound to the way of the traditional instrument, the easier can perceive the recipients the interactive system.

But about the <interaction 2>, that is the interaction between the gesture and the image, composers should answer with imaginative interactivity including double visualization.

To get references about this problem, I take three examples with musical actions. The two have made answers with presenting the coordination of contents of the image to the total sound and musical contexts; one is taking such coordination as in the dramatic theater or

music theater, and another is taking the motions of the gestures as the coordinate from sound to image. The former postulates the common cultural background to the creator and to the recipients, because each culture has the original body action expressions. The latter appeals directly to the recipients' audio-visual perception regardless of their cultural background.

The third example has made answer with an impressive interactive live stage in the way that two different audio-visual spaces are presented parallel and that the interaction can be perceived depending upon the aesthetically synthetic treatments inside the recipient.

Emphasized Gesture of the Instrumentalist as a Performer of <Musiktheater>

Shigenobu Nakamura has about twenty-year history of the composition of <music theater>. The term has come from Musiktheater of Mauricio Kagel. Nakamura's audio-visual interactive pieces are conceived as theatrical pieces, where the gestures of the performers have the theatrical meaning besides the musical meaning. Especially the audience cannot help rethinking the actions of the instrumentalist because the camera makes emphasis on his action. And the images projected on the stage are functionalized as visual *objet d'art* on the scene³.

Nakamura's <Kagami>(1996) is composed for a player of trumpet, cameras and the interactive computing system. The player's actions are off course for the sake of musical performance, but at the same time the actions can control the switching from one camera to another. For the recipients the interaction system is sometimes invisible but they perceive the relationships between the actions and the images thanks to the contents of the images, that is, the player's realtime actions. The speed of the graphical processing makes the movie images unify to the performer's actions.

Gesture of Instrumentalist as a Direct Sound/Image Generator

Shu Matsuda is the key person for the development of DIPS program⁴ based on jMax(2). In <Force Feedback, for flute and computer> (2003) realized by Shu Matsuda with DIPS, the captured sound and motions of the instrumentalist are reflected on the sound and the three dimensional video effect. The impressive 3-D effects operated on the realtime camera-captured video, and thanks to the musical context played by the instrumentalist and the fast graphical processing the interaction can be perceived directly without recurring to the meaning of the gesture.

Gesture of Instrumentalist as a Mediator between the Two Heterogenous Audio-visual Spaces

Masayuki Akamatsu is famous for the Max tutorial textbook <TransMax Express>, DSP Summer School and DSP magazine. The Akamatsu's audio-visual interactive piece <mass & media>(2004) is called as percussion performance with TV broadcasts. This is the piece for a percussionist and the live-computing with videonics on Max.

In this piece, the percussionist plays with the traditional notation. His actions are functional as musical gesture and at the same time operate as instructions to the computer program which controls the play speed of each frame and the change of the live TV channel. Each frame is the result of thirty partitions per second.

During the performance, the recipients can see on the stage the percussionist's gesture, the screen where is projected the controlled TV program, and the monitor which shows the information flow. The recipients experience two different audio-visual spaces at the same time. One is the space where the percussionist's gesture relates the percussion sound as in the traditional performance. The other is the TV program's audio-visual space, which are popular to the Japanese habitants. The two spaces are definitely different on the point of audio-visual contents even though each of the two is very popular. So the parallel proceeding of the two different audio-visual causality is perceived very clearly by the recipients. And the interaction between two spaces with the special system can be perceived through the common motion of each audio-visual causality; the motions like stop, start and the interval or the speed of stop/restart etc. can be experienced parallel in the two audio-visual space in each corresponding moment.

As the final examples, I present some pieces with non-musical actions. Gestures except for musical action, especially the motions of dancers, are often used in interactive pieces. I take two examples; one is with visible actions and another is with invisible actions.

Nakamura's <Shadow of RAMA>(2004) has the interactivity with which the moving hands of the performer(=composer) controls the shape and color of the image. The gesture of the performer generates no analogue sound and the image is simple, so the recipients can gaze at performer's motions and the screen. They concentrate their attention on the interactivity, so they easily find out that one action of the performers lead to one musical event and one video processing event. In this case the gesture is drama action.

Laptop-based computer manipulation in music performance is more problematic.

In Kiyoshi Furukawa's <In the shape of green >(2003), the image on the screen operated by the performer(composer) is the reason and the sound is the result. The performer's operation is like a kind of visual game to make some moving shape with colorful balls and lines. The performer responds to the sound as the result of the visual game, and decide the next motion like improvisation. Here the interaction is cyclic from visual control, sound as a result, and to visual control as a human response to the sound.

Nao Tokui's live-performance with SONOSPHERE is also takes in the non-musical information. The program <SONOSPHERE>is a system for visualizing of music/sound process, which depicts the functional units of musical processing as round shape form in the virtual space(5). The round shape forms can be connected with each other as in MAX. TOKUI told decisively the laptop music as a media art from the traditional music with musical instruments and the performers. Depending on him, the laptop music is not to be supposed to postulate the player's virtuosity and the cultural background of the interpreters. With his live performance the recipients see the explanation of the sound processing.

My audio-visual interactive piece <quantumcumque> is an example with today's MAX for audiovisual interactivity where non-musical action generates sounds and images. This piece is composed for a flutist, a drawing performer and a computer manipulation. It includes a system which realizes the exchange the pen tablet information for the sound and video effect. The system configuration is in [Figure.2].

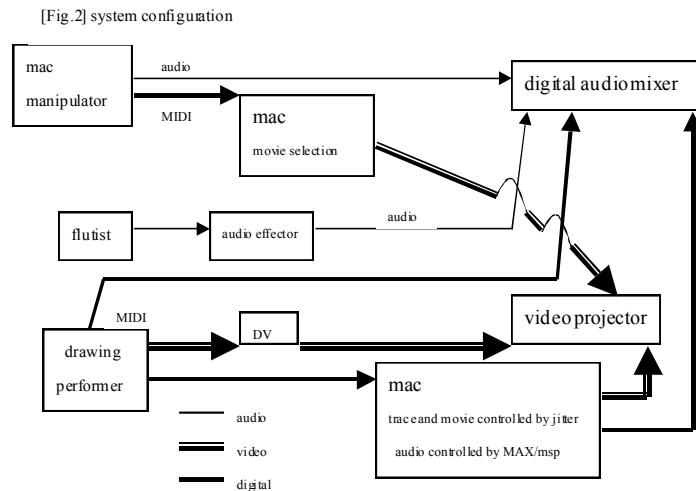


Figure 2. Configuration of <quantumcumque>

The main technical interaction here is that tablet input is affecting on the sound and the screen. The image is consisted of three elements ; the drawing performer's realtime motions through the camera on stage, the drawn sketches which are the traces of the pen tablet and also some movie files including graphics relating to the contents of the drawn sketches. The flutist has a traditional ensemble with electronic sounds. The relation between the drawing man and the flute and electronic sounds is a kind of human communication. The drawing performer thinks that both the flute sound and the electronic sound are equivalent in the point that both are the music. Similarly seems to the recipients in the concert hall.

There are three kind of interactivities; between flutist and electronic sound, music and movies(audio-visual), drawing performer and movies. Only the last interaction has technical system(from pen tablet to movies). All the others are interactive only in human communication.

The time structure are led by the flutist, so is monolithic and simple. But it is not linear. The three kind of human interactivities make natural ensemble character as those in the case of two or more musicians play together, and these interactivities are in different dimension from the audio-visual interaction for the audience. So the audience level of time structure seems to be independent of the technical interaction. Human interaction in the performing level can make a linear time structure, but pen tablet system and the recorded electronic sounds make the linear time multiplied with the contradictory time.

In Chikashi Miyama's <Babel>(2004) , mixture is realized in more sophisticated way. This piece includes DIPS and two voice performers. The interaction can be perceptible directly like in the case of Matsuda. Miyama has constructed theatrical five scenes, each of which has different priority on music, visualization and drama. So the interactivity has characterized differently to each of the five parts. Even though the recipients can see the interaction but the interaction is not the main point of the piece's visual stage design and the duration of the piece is long enough, so the screen plays the roll of main visual events of the piece.

Multiplied interaction

Today's audio-visual musical works with interactive system hold two levels of interaction: traditional musical ensemble interaction, and, man-machine interaction. Traditional musical ensemble has both audio and visual elements concerning generation of sounds, so in the case of audio-visual interactive pieces, the causality of sound generation is blurred, while the causality of image generation is clarified. But by overlapping the real-time performing movie on the processed images the situation is the same as that of sound generation. The composer has multiple possibilities to make various type of sound-image relation by selecting sound/image file, overlap/non-overlap/structured relationship between more than one materials, and so on.

In this presentation *Gesture* is represented as motions of musical performance like playing the traditional or newly developed instruments or singing/recitation. But in such pieces as <Small Fish> by Kiyoshi Furukawa and Masaki Fujihata and such live performance as with <SONASPHERE> by Nao Tokui the presentation form/style is absolutely different from that of stage-audience musical presentation. Both performances use lap top computer as controller to lead gestures.

SONOSPHERE is a system for visualizing of music/sound process, which depicts the functional units of musical processing as round shape form in the virtual space. The round shape forms can be connected with each other as in MAX. Tokui distinguishes decisively the laptop music as a media art from the traditional music with musical instruments and the performers. His proposition is based on the viewpoint of interface. He asserts that the traditional music or the software for that kind of music depends on two interfaces; score or notation and traditional hardware like musical traditional instruments. Both of the two postulates virtuosity and the cultural background of the interpreters. TOKUI demands another type of interface for laptop performance.

One important point is that <Small Fish> is the name of the piece and that <SONASPHERE> is the name of the system. Their live performances are actually live-computing musical practices, but the concept such as <piece> or <work> of each performance are not essential. They are quasi-improvisation with newly developed sound-image system. So the aesthetic roll of gesture in audio-visual interactive musical piece in those performances are decisively different from those of on-stage pieces.

Today the term <interactive> is global beyond the genre of arts and beyond frame of social functions. So the discussion of models for interactivity should include such basic issues as what is music, what is the aesthetic essence of the performance or which relationships the piece has to have to which kind of music/art, etc. It is desirable that analyzing of the situation like this paper should operate as mediation between music history and music computing performance.

¹ Formant Brothers. <http://www.bekkoame.ne.jp/~nsakonda>

² Akamatsu, Masayuki. <http://www.iamas.ac.jp/~aka>

³ Nakamura, Shigenobu. The Possibility of Musical Representation with Visual Elements. In: IPSJ SIG Technical Reports 54.

⁴ Matsuda, Shu., Rai, Takayuki. A Visual-to-Sound Interactive Computer Performance System Edge. In: Proceedings of the International Computer Music Conference. 1995.