Robotic Interface for Embodied Interaction via Dance and Musical Performance

KENJI SUZUKI, MEMBER, IEEE, AND SHUJI HASHIMOTO, MEMBER, IEEE

Invited Paper

A substantial robotic interface is proposed for collaborative work between humans and machines in a multimodal musical environment. The robotic interface is regarded as a "moving instrument" that displays the reactive motion on a stage while producing sound and music by embedded stereo speakers according to the context of the performance.

In this paper, we introduce four musical platforms utilizing robotic technology and information technology in different circumstances. These are effective designing environments for artists such as musicians, composers, and choreographers, not only for music creation but also for media coordination including motion and visual effects. The architecture, called the MIDI network, enables them to control the robot movement as well as to compose music.

Each of the developed robotic systems works as a sort of reflector to create an acoustic and visual space with multimodality. The proposed approach to equip musical instruments with an autonomous mobile ability promises a new type of computer music performance.

Keywords*—Computer music, human–computer interaction (HCI), hyperinstruments, multimodal human–machine interaction, robotic interface, user interface human factors.*

I. INTRODUCTION

Music is a typical channel of nonverbal communication [1] that humans often use to express their feelings. It is essentially a mode of expression and a mode of emotion, affection, and also a Kansei communication channel among people. *Kansei* is a Japanese word that means something like "sensitivity," "intuitiveness," and "feeling." People have certain feelings about the use of the word *Kansei*, although it is not measured in a quantitative way, nor is it visible like the feelings of people [2]. We consider that a musical performance realizes a Kansei interaction to express individual ideas and thoughts with the aid of instruments.

We have been investigating and exploring the paradigms of the embodied interaction between humans and a robot in the framework of museum exhibitions, theatre, music, and art installations. Although humans often accompany music with body motion, few studies have reported the autonomous mobile robot for a musical performance. As Pressing [3] remarked about musical instruments from an ergonomic point of view, the design incorporated not only a particular musical function but also the type of feedback, i.e., tactile, kinesthetic and visual feedback. Also, Bahn *et al.* [4] focused on the human body itself in the musical performance and addressed that musical performance has been inextricably linked to the human body.

Consequently, focusing on an interaction metaphor that a robotic interface establishes a virtual and real-world connection, we have proposed some sensory systems and mobile robot platforms for the multimodal musical environment. The first idea was to equip a musical instrument with an autonomous mobile ability for computer music performance. The robot can be effectively used for musical performances with motion because it can move around while generating sound and music according to the performer's movement, environmental sound, and image. We are interested in the robot that displays the reactive motion according to the context of the performance to create the human–robot collaborative performance on a stage. The developed system is regarded a "moving instrument" and is a sort of active aid for a musical performance that allows the users to get feedback for emotional stimuli in terms of sound, music, image and motion. Applications for interactive art, music, edutainment (education and entertainment), and the rehabilitation therapy (e.g., of autism [5], [6]) are promising fields of this study.

In this paper, we describe a comprehensive study of robotic interface for dance and musical performance. We

Manuscript received February 2, 2003; revised October 27, 2003. This work was supported in part by the 21st Century Center of Excellence (COE) Program "The innovative research on symbiosis technologies for human and robots in the elderly dominated society," Waseda University, Tokyo, Japan, in part by the JSPS Research for the Future Program in the Area of Kansei (Intuitive and Affective) Human Interfaces, and in part by the Japanese Ministry of Education, Science, Sports and Culture under Grant-in-Aid A-11305021.

The authors are with the Department of Applied Physics, Waseda University, Tokyo 169-8555, Japan (e-mail: kenji@ieee.org; shuji@waseda.jp).

Digital Object Identifier 10.1109/JPROC.2004.825886

first describe the background and related works. Four case studies will follow the modeling of the human–machine environment in music-based interaction of Section II-A. Section III presents an environment-oriented interaction, namely, a reactive audiovisual environment. Sections IV and V present two types of semiautonomous robotic interfaces, Visitor Robot and the iDance platform. Section VI describes an autonomous robotic interface, MIDItro. The concept of the Musical Instrument Digital Interface (MIDI) network and a discussion are then given in Section VII. Finally, a conclusion and future works will be discussed in Section VIII.

II. Modeling of the Human–Machine Environment

A. Background and Related Works

The computer-generated music have been the focal point of computer musicians and scientists for many years, e.g., [7]–[9]. For increasing the degrees of freedom of musical expression, a number of researchers and musicians have explored new types of musical systems that overcome the physical limitations of musical instruments and the human vocal cords [10], [11].

These musical systems can be divided into two categories. The first category features a new kind of musical instrument to enhance expressiveness by utilizing digital signal processing for adding sound effects based on traditional musical instruments. Live interactive electronic music by hyperinstruments based on the organ, percussion, and cello have been performed in the early stage of these studies [12]–[14]. More recently, Cook *et al.* [15], [16] have developed an accordion that embeds a microcomputer for controlling the sound. A Japanese traditional instrument, the Sho, with attached sensors to sense the breath of the player has been developed by Nagashima [17].

The second category features multimedia systems that utilize haptic, video, and motion capture devices. These systems differ from traditional musical instruments in the shape and the interface. The theremin is the first electronic instrument that is played without being touched. The gesture of the player can control the pitch and volume of the sound. There have been a number of attempts to create music according to human gesture and entire body movement [18]–[21]. Various types of sensing techniques have been used to detect human motion, and the measured body movements are mapped to music or sound. Especially, a video-to-music system has been emphasized with the aim of investigating to associate natural gestures with sound/musical features for interactive performance, for example, [22]–[24]. With a particular focus on facial expression, Lyons *et al.* [25] have developed a musical controller that converts the user's facial feature parameters extracted from live video input to expressive musical effects. As for a haptic interface, Sawada *et al.* [26] introduced a ball-shaped instrument that can sense human grasping and the movement. One of the aesthetic interactive devices is the interactive kaleidoscope, called Iamascope [27], which incorporates graphics, vision, and audio technology to create sound and imagery in an interactive device.

Some researchers and artists have emphasized to construct an interactive musical environment. As an example of such an environment-oriented musical system, Rokeby [28] has attempted to construct a space called the *Very Nervous System*. It is an interactive sound installation in which the movements of one's body create sound and/or music by utilizing multiple video cameras. The system has been presented as an art installation in galleries and public outdoor spaces, and has been used in a number of performances. The *Virtual Cage* by Möller [29] is a floor for music creation which can sense the weight shift of a human on it and can create acoustic feedback. At the same time, the pneumatically controlled platform can move and give physical feedback to the user. The *Intelligent Stage* [30] is also a typical media-controlled space equipped with multiple video cameras and video projectors, which trace and projects a performer's action within a structured environment. The environmental reaction can effect the environment itself in accordance with the design of the composer, choreographer, or director.

In recent years, the human–machine communication by means of speech, gestures, and haptics has been implemented in various scenes of our lives with the aid of the growth of multimedia technology. For example, pet-type robots have become commercially available by a number of industrial companies. They can exhibit attractive and devoted behavior so that the audience is satisfied with its performance. These are examples of an advanced interface that has a substantial body with multimodality. Such multimodal human–machine interactions typified by nonverbal communication have been widely investigated in different contexts; for example, some approaches to build a socialable robot have been reported in [31] and [32].

Some attempts have been reported a mobile robot in order to connect the virtual and real musical worlds. Eto [33] produced a network-based robotic art installation. The robots in the site for exhibition communicate with each other and create music. These robots are also controlled by users at the site or via the Internet. Wasserman *et al.* [34] reported a robotic interface for the composition system called *RoBoser*. It is a small mobile robot that autonomously interacts with its surrounding environment and has an algorithmic computer-based composition system.

B. Modeling of Human–Machine Environment

In a multimodal musical environment involving a robot, we consider that the style of interaction can be modeled along two axes as illustrated in Fig. 1.

The horizontal axis represents the autonomy or embodiment of the instrument. The system in the positive direction intends a stand-alone or elementary substance that is denoted an object-oriented system, while the system in the negative direction represents a musical installation and system circumstance that is denoted an environment-oriented system. The vertical axis represents the degree of the direct/indirect

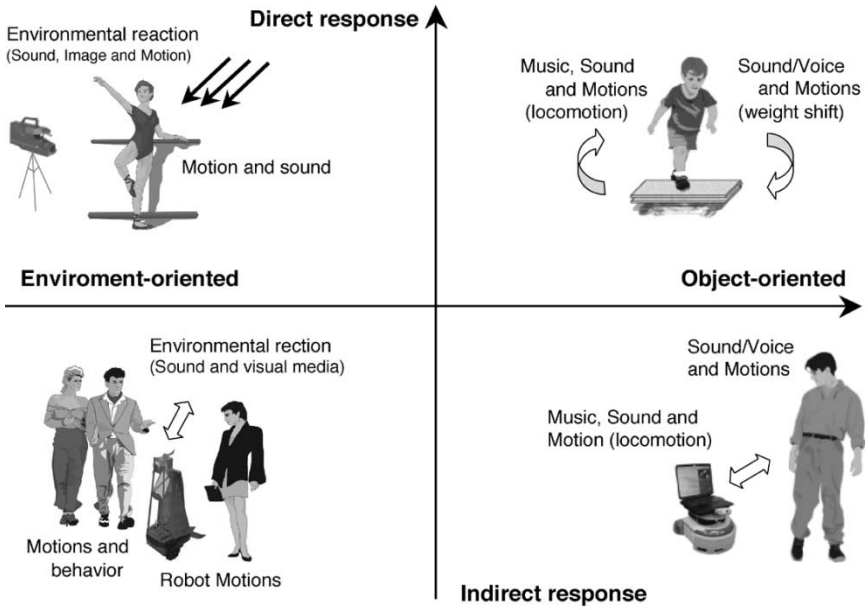


Fig. 1. Modeling of human–machine environment. The style of the interaction can be modeled along two axes: the autonomy and embodiment of the robot and direct/indirect response. The region is largely divided into four areas according to the characteristics.

reaction. The system in the positive direction directly responds in accordance with the external environment, which features a kind of a media converter or reactive systems, while the system in the negative direction does not respond directly to the environment, but provides its output that is determined from the environment and its internal state.

The systems having mobility are categorized into the following four types. The description of each area is as follows.

1) Left-upper quadrant of Fig. 1: Reaction of robotic environment by means of music/sound, image, and motion is caused by the emitted sound and human behavior in the environment. We consider this a model of an environmental robot. An environment is "robotized," which means that people interact with the surrounding environment where human motion and the emitted sound can be sensed by the environmental elements such as wall, objects, and room. The surrounding environment then responds by generating sounds, images, and motions. Although the environmental motion is not included in the present study, further consideration includes the phenomenon that the wall moves to people, or an object transforms according to the stimuli.

   The video-to-music works and environment-oriented systems such as the *Very Nervous System* and the *Intelligent Stage* are categorized in this domain.
2) Left-lower quadrant of Fig. 1: The robot makes its own motion and displays music/sound and image from the surrounding environment, in accordance with environmental sound and image surrounding the robot. The robot plays a role of an advanced interface, and its reaction, including music and visual effects, is produced by environmental elements. The *RoBoser* is an example of this domain.
3) Right-upper quadrant of Fig. 1: Through direct physical contact between human and robot, the robot follows the applied force and also displays its own motion and creates sound and music according to the environmental sound and image, and an applied force caused by physical contact. Note that the robot is not only an interface but also a sound production device. In this sense, conventional musical instruments and some hyperinstruments (for example, [35]) are categorized in this domain.
4) Right-lower quadrant of Fig. 1: Through direct/indirect contact between a human and robot, the robot autonomously displays its motion and creates sound and music according to its internal state and the environmental sound and image. No existing system or approach, as far as we know, attempts to construct an autonomous musical robot.

We have realized the above four systems typified by each category for music-based human–machine interactions. In these frameworks, we made a condition that humans do not need to wear any special on-body sensors for the flexible performance.

Regarding to the hardware architecture, the developed systems have realized the basic concept of the MIDI network, which provides a seamless communication among the devices of the proposed moving instruments. MIDI is a digital language and a sufficient communication protocol, which has been accepted with industry professionals, computer music composers, artists, and major manufacturers of electronic musical instruments for many years. In this framework, MIDI is used not only for sound and music creation but also for a protocol of motion command to the robot. The details of the MIDI network will be described in Section VII.

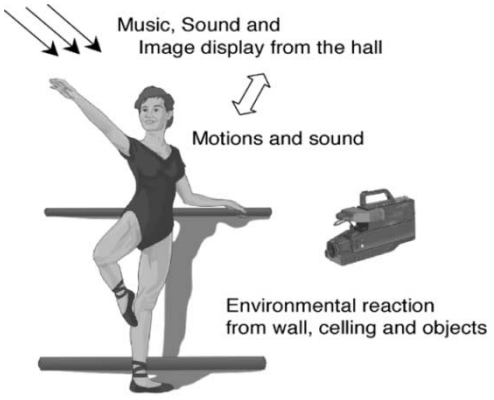


Fig. 2. Case study I: reactive audiovisual environment, which is a style of music-based human–environment interaction [36].

Regarding to the software architecture, on the other hand, all components of the sound analyzer and synthesizer, image analyzer, and behavior coordinator are constructed with Max/MSP modules. Max/MSP is widely known by composers of computer music. It is an object-oriented programming environment with a powerful graphical user interface to deal with MIDI data. The system designers, therefore, can easily treat the operations and relations by changing the connections among these Max/MSP modules. In this paper, *patcher* means a collection of modules in the Max/MSP environment.

III. Case Study I: Reactive Audiovisual Environment

A. System Overview

In this section, we describe an advanced interactive system, which can create sound and music in real time by automatically extracting structural features from a moving image and environmental sound, and associating them with musical features. This is a case study of an environment-oriented musical system as illustrated in Fig. 2. The movement of a performer can be acquired by a video camera and a microphone installed in an interactive space, where the performer freely dances and emits a voice or handclaps. The system responds according to the system design of a composer or a musician by means of sound, music from the installed loudspeakers, and image by a video projector. The system, thus, provides an interactive musical environment.

The system is operated in the Max/MSP environment on a Macintosh G3 without any other special equipment. Fig. 3 shows a description of the developed system consisting of two modules (*image analyzer* and *sound analyzer*) and a patcher (*music generator*). First, moving images are captured from the charge-coupled device (CCD) camera at the rate of 10 frames/s. The image analyzer calculates the color information, and the spatial and temporal features of the image. At the same time, the environmental sounds are captured from the microphone attached to the system. The sound analyzer then calculates the auditory information of the sound. Finally, the music generator tunes the chord progress, rhythm, and tempo by associating the image and sound features with the musical features and creates the music.

B. Sensing the External Environment

1) Image Analyzer Module: The *image analyzer* module can extract temporal and spatial features [37] from moving images. The input source of the module is a moving image from the video camera, while the output consists of image features; color information such as red, green, and blue (RGB) and hue, saturation, and lightness (HSL) components, and the spatial and temporal features such as the density of the edge, pattern data, blinking information, and scene changing value (binary data).

In the present study, the size of the captured frames are 320×240 pixels. Each frame is divided into $M \times N$ areas (in the present study, we use $M = N = 3$). The features of the moving image are calculated in both the whole frame and the image of each area to obtain the global features and the local ones. The summation of the values of RGB and HSL components obtained by the image data in each frame is used for the spatial features. By using this color information, the average values of the edge density are also extracted as the spatial features by a two-dimensional (2-D) filter F_{edge} as

$$F_{\text{edge}} = \begin{pmatrix} -1 & 0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{pmatrix}. \tag{1}$$

On the other hand, for the acquirement of temporal features, the image analyzer stores basic frame data as a background image at the beginning of the detection of the moving image. By comparing to the background image, scene changing can be detected by calculating the temporal difference for every newest frame

$$\begin{cases} D_k(p,q,t) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |u_k(t)(x_i,y_j) - u_k(0)(x_i,y_j)| \\ D_{\text{all}}(t) = \sum_{k=H,S,L} \sum_{p=1}^{M} \sum_{q=1}^{N} D_k(p,q,t) \end{cases} \tag{2}$$

where $D_k(p,q,t)$ denotes the difference in brightness at each divided area (p, q) and time t. m and n represent the width and height of each divided area. $u_k(0)(x_i,y_j)$ and $u_k(t)(x_i,y_j)$ represent the value k $(= H,S,L)$ of the background and present image at a pixel (x_i,y_j), respectively. An example of the extraction of scene changing is illustrated in the left-bottom of Fig. 4, where the x axis represents time t, while the y axis represents D_{all}. The threshold θ represents the horizontal line in the figure. When D_{all} exceeds the threshold, a scene change has occurred.

2) Sound Analyzer Module: The component called the *sound analyzer* works to extract the sound features and auditory information of the environmental sound. The input source of the module is the sound wave from the microphone that comes equipped with the standard Macintosh MIC-in. This component calculates pitch and amplitude as the sound features and performs auditory scene analysis including the estimated environmental state and the scene change.

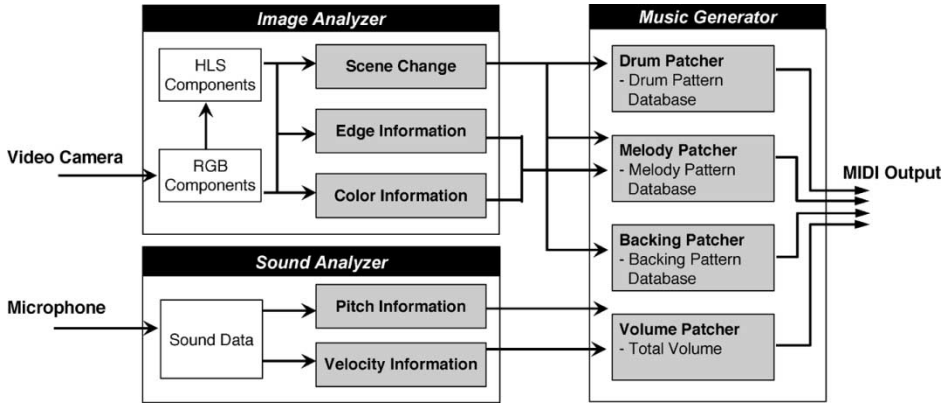


Fig. 3. System overview. The system consists of two modules (*image analyzer* and *sound analyzer*) and a patcher (*music generator*).

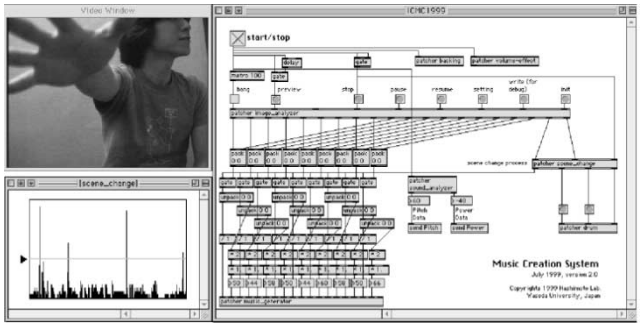


Fig. 4. Sound/image analyzer patcher on Max/MSP environment. The top-left figure is a camera window to display the captured image. The bottom-left figure shows the value for the scene changing.

In this study, we assume that the given sound is a single source, for example, a voice of the performer, a whistle, or a handclap. The sound volume from each microphone is obtained by a Max/MSP module. The cepstrum method is then adopted for the identification of the fundamental frequency. The method uses the harmonic structure obtained by Fourier transform at the high range of the cepstrums. In order to decrease sampling errors, the fundamental frequency is regarded as a value of the nth peak of the frequency divided by n. In addition, the system can also recognize the state of the environment from the auditory information. We equipped with a programmable composition function to change of sound amplitude and spectrum density in order to distinguish each mode of auditory information such as noisy, silent, and singing.

C. Environment Reaction: Music Generator

The patcher called the *music generator* creates the melody, chords, and rhythm. This component can associate the image and sound features extracted from the above two analyzer modules with the global musical features (such as the chord set, rhythm, tempo, timbre, and volume). This patcher consists of the following internal modules: backing chord progress generator, drum pattern generator, melody generator, and volume generator.

The correspondence of the image/sound features with musical features can be predetermined or interactively modified by the user. The music generator accepts data not only from the above two analyzers, but also from the external signals such as tuning the parameters on this Max/MSP module. The internal data set of the chord patterns such as backing scales and chord progress is involved in the music generator. Although the basic rule of composition is adapted in advance based on musical theory, the user can add new rules for the specific algorithmic composition. The music generator then compiles all the determined musical features into the MIDI sequence. The modifiable features include pitch (note number) and pitch-bend (to change a note's pitch in a continuous sliding manner), timbre (choice of instruments), volume, tempo, rhythm pattern, chord progress, and melody line.

In the present study, the melody notes are created according to changes in the color information, while the MIDI velocity of each note is produced by the edge density. A MIDI note on event consists of pitch (note number), velocity, MIDI channel (timbre), and time. MIDI velocity is regarded as a key pressure.

In addition, by using the key information such as scene changing, the system changes the current mode of the music generation. Two examples of music generation mode are described as follows.

1) Rule-Based Music: We humans are familiar with the created music based on musical theory. Also in the field of computer music, the structure of chord progress and melody harmonizing has been often applied from many kinds of musical theory. In this study, backing chord progress and the drum pattern are renewed every eight bars of generated music, or the instant of a scene change. The backing chord is selected from five typical prepared patterns. In addition, the base note of the backing chord can be changed with seven steps of key notes defined by the obtained pitch of the auditory information. With regard to the drum pattern, 72 sets that have six different tempos are available. According to the change in the temporal features of the image, the rhythm of the created music changes with time. On the other hand, the kinds of instruments are also changed at the instant of a scene change.

Fig. 5. Case study I: reactive audiovisual environment. Two dancers freely perform in front of the equipped CCD camera. The obtained moving image is displayed on the screen of the wall.

2) Stochastic Music: Another kind of music generation, namely, the stochastic music mode, is installed. Up to now, several research studies about stochastic music generation have been reported; for example, some classic works are found in [38] and [39]. In this study, every beginning phrase of starting this mode, the note set is determined by the input data from the sound/image analysis. The chord progress and melody can be created with a random value within the note sets. The prepared four kinds of note sets are major pentatonic, minor pentatonic, Japanese major scale, and Japanese minor scale.

D. Interactive Music Generation

This system starts to generate music when an object appears in front of the camera or when the microphone captures an environmental sound. In this section, we describe several kinds of sound effects between the user and the system.

1) Scene Changing: The intense scene changing defined as temporal feature changes cause the renewal of all the musical features such as the choice of instruments, backing chord, drum pattern, and rhythm.

2) Volume Modification: According to the change in the sound volume obtained by the sound analyzer, the total volume of the generated music/sound is modified. In order that the system follows a human's vocal performance, the change in volume is done with a short time delay.

3) Stereo Sound Field: When an object moves in the right/center/left area of the captured image, the relative level of the created sound is fed into the respective left and right loudspeakers according to the location of the moving object. In addition, each of the three areas has a different pattern of music creation. Therefore, the system allows the user to express his intention as if he has three kinds of instruments.

4) Harmony Generation: When a user starts to sing in front of the microphone, the system generates a harmonic sound. With a short time delay, a single note of piano sound is produced so that the system would harmonize with the captured voice or whistle.

E. Performance Demonstration

We have done a dance performance with two performers, as illustrated in Fig. 5. The developed system is installed in a room where one video camera and a microphone were placed. The created music from the environmental and auditory information is produced in real time from the stereo speakers. The obtained moving image is displayed on the screen of the wall.

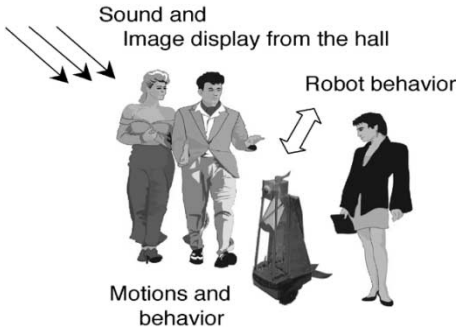


Fig. 6. Case study II: Visitor Robot, a semiautonomous robot for music-based human–machine interaction [41].

The musical features such as timbre, pitch, volume, tempo, and style of music generation are modified by the system designer in real time. The variation in music generation and the interactivity was significantly shown in this demonstration (see also [40]).

IV. Case Study II: Visitor Robot

In the another case study of the environment-oriented robotic interface, we introduce an intelligent agent for interaction between humans and robots embedding artificial emotions. Unlike the system directly responding to the external environment, the gestures of a human, scenes, and sounds in the surrounding environment can cause the robot to change its internal state, its behaviors, and its reactions within the interactive space. The robot, thus, responds not only to the given stimuli but also reflects its internal state in accordance with the design of the system designer.

The developed robot freely tours in the exhibition as one of the many people who frequent it, a sort of medium between humans and machines living together in the exhibition area as illustrated in Fig. 6. It exhibits its current emotional state by means of integrated visual media, environmental lights, music, and changes in behaviors and style of movement.

A robot embeds a computational model of artificial emotion, which is constructed by taking advantage of the self-organization of an improved model of Kohonen's *Self-Organizing Map* [42]. The network is adapted so that it can represent an emotional state; the current emotional state of the robot is determined as a result of competition with

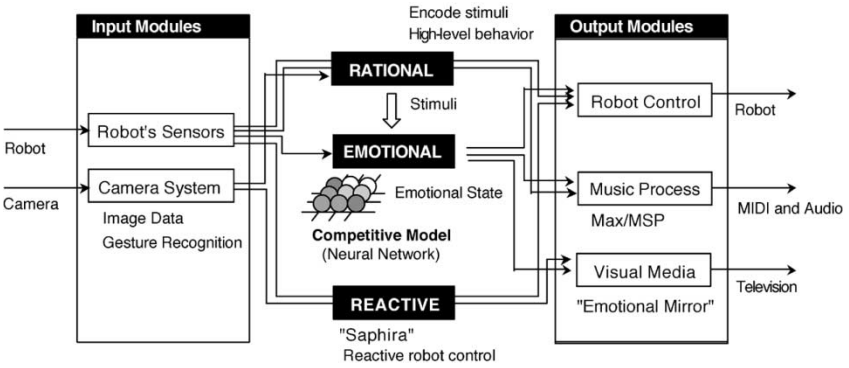


Fig. 7. An emotional agent architecture *Robotic Agent* for multimodal environment; the details of the agent architecture are described in [43].

other states modeled by the network. Moreover, the state changes dynamically and represents the "personality" of the robotic agent.

A. System Overview

The small robot on wheels is a Pioneer 1 robotic platform by ActivMedia, Inc., which is equipped with an onboard camera, infrared localization sensors, a local audio system, and two wireless communication channels for both audio and video signals. The inputs of the robotic agent are given from the robot's low-level sensor data and the onboard camera. As outputs, the system integrates three kinds of communication channel: movement (the behaviors of robot), visual (environmental light), and acoustic (music and sound).

Fig. 7 shows the dataflow detail in the developed emotional agent. The outputs are produced through three components: rational, emotional, and reactive. Each module operates under the influence of the other modules. As a whole, they process input parameters from the external world and produce output parameters. Note that only high-level information is processed through the rational and emotional modules. The *rational* produces controlling data of the behavior of the robot and stimuli to the emotional module in order to drive the artificial emotion. On the other hand, the *reactive* module produces parameters for dynamic output. The *emotional* module is the core, which changes the current emotional state of the robotic agent. The emotional state, which represents the personality of the robotic agent, then influences each communication module at the *output* component.

The system for the robotic agent works under the Win32s operating systems. The robot communicates by three different radio links [digital input–output (I/O) control data, video, and audio signals] with the supervisor computer on which the model of artificial emotions is realized. The robot also possesses an onboard audio diffusion system, connected by radio, which integrates the audio diffusion system placed in the environment. Three computers connected by an Ethernet network and the MIDI network control different various aspects; the first contains the emotional model and control movement, the second deals with "emotional mirrors," and the third generates sounds and music.

B. Sensing the External Environment

1) Sensor Input: Input module consists of two main components. One is a receiver from the robot's sensor. Its role is to gather data from the robot's sensor every 100 ms. The following data are computed: the absolute and relative positions of an object and the robot, and logarithmic distance and area when an object is near the robot. The space around the robot is divided into five areas that are observed by five ultrasound sensors, placed in front of the robot. The sensor data gathered by *Saphira* [44] is processed by the *input* module. *Saphira* is the special purpose robotic software that was developed by Stanford Research Institute (SRI), Stanford, CA, in order to handle the low-level details of the robot, such as drive motors and wheels. In other words, the component observes the environment around the physical relationship between the human and the robot.

2) Gesture Recognition System: Another component is a camera-based sensor system, which allows human gestures as inputs of high-level information. This system allows a human to communicate with the robot with the aid of a small light source.

The detected position of the light is sent to the robotic agent about 20 times per s. The detection starts when the user turns on the light in front of the robot and ends when he turns it off. This duration is determined as one phase. Each gesture is normalized to the smallest rectangle from the center of the light trace. It, therefore, does not depend upon the area of the human's gesture. The detected area is then quantified into an 8×8 image that is used as input data for the recognition processes. The agent extracts not only the pattern of gesture, but also the size of the detected area and duration of one phase.

A low-level processing and a simple back propagation neural network are used for the gesture recognition. In the present study, the data is sent to the robotic agent as "negative" and "positive" information for changing the emotional state. This recognition result is not used for real-time musical interaction nor a reaction of the robot because the latency is quite slow for these purposes. Ten gestures seem to be sufficient for communication with the robot. The circle gesture, for example, can mean a positive stimulus, while a slash gesture can mean a negative one. The time of recognition is less than 1 s.

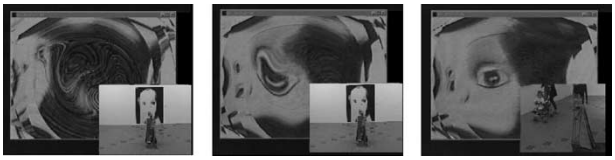


Fig. 8. Dynamics of emotional mirror. What the robot sees appears "mirrored," corresponding to the positive and negative emotional states. During the performance, the system shows such images on a TV screen in real time.

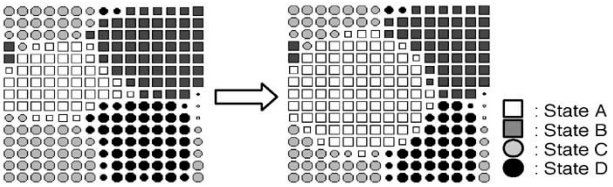


Fig. 9. The change in the emotional state after a stimulus is given. The model is inspired by the dynamics of human emotion.

C. Robot Reaction

1) Robot Control: Output module consists of three components. One is a component to control the robot. Two types of control are prepared: behaviors and movements. The former means behaviors at a high level so that the robot might produce performance similar to the human. In short, it seems that each behavior has a reason, such as following a human or escaping, attention or avoidance, or turning around a human. On the other hand, the latter corresponds to quite simple movements: forward, backward, and turns. These types of behavior and movement patterns are controlled by orders from the *rational* component.

2) Music Generation: The robot agent outputs particular data to the Max/MSP so that the generated music can reflect the emotional state. In the Max/MSP module of music generation, the data are arranged so that it can reflect the emotional state and style of movement of the robot. The robotic agent generates parameters used to modulate the score skeleton in each time slot. They include not only the emotional state but also the movement of the robot through its *reactive* module. They consist of emotional parameters and physical relations between the human and robot such as distance and area. For example, the music volume is changed by the physical distance obtained at the *input* component between the robot and human. In addition, when the robotic agent receives gesture information, parameters are also sent to the music process, thus influencing the score skeleton.

3) Visual Media—The "Emotional Mirror": A further output is the control of visual media. It is available to show the emotional state of the robot, and also the user can see what the robot looks at. It is expected that the users can understand the state of a robot more clearly with different output channels.

The visual component is based on the idea of an "emotional mirror." Fig. 8 shows an example of the aspect and the implementation of the emotional mirror. Referring to the related work, the *magic morphin mirror* by Darrell *et al.* [45] provides a self-referential experience by a virtual mirror with face-specific image manipulation. The system demonstrated the morphing and texture synthesis for interactive displays as a result of a face tracking and expression analysis.

In this study, the concept of the magic morphin mirror is extended so as to enable people to understand the internal state of the robot with a modified image captured by the onboard camera. The robot sees what is in front of it (people's faces, artworks) warped according to its emotional state. For example, a human or a face could appear "mirrored," distorted in a vortex or reprocessed with bright colors, respectively corresponding to emotional states. During the performance, the system shows such images on a TV screen in real time.

D. A Model of Artificial Emotion

One of the motivations to construct artificial emotion is the complexity while making decisions in the robotic agent. The artificial emotion is a heuristic model that is one of the means that supports the agent making decisions dynamically and flexibly. The robotic agent divides given inputs into only four vectors for simplicity. The artificial emotion model then consists of the corresponding four states, which form the personality of the robotic agent. This structure is called "emotional space," as shown in Fig. 9.

The model is composed of four characteristics that represent the emotional state, and each symbol corresponds to a particular state of emotion. Each state with a typical human emotional condition is called *happy*, *angry*, *melancholy*, or *tranquil* for simplicity. The number of each symbol (cell) represents the rate in emotional space, and each state represents a unique character of the robot.

Considering changes in the emotional state as a result of the competition with different emotional states, an emotional model was constructed based on a self-organization map (SOM) [42]. The network is improved so that it is suitable for dynamic changes in the emotional state. The architecture of the modified self-organization map consists of 15×15 cells in 2-D with a torus structure. Each cell has dimensions of emotional states. Here, the input space is five dimensions that correspond to the estimated relative position of an object or a human around the robot and the result of the gesture recognition. When a human gives information to the robot, the agent can understand it by its strength. This parameter corresponds to the learning parameter of the network. The larger this parameter is, the stronger the influence on learning in the network (see details in [41]).

The change in the emotional state is shown in Fig. 9. The kinds of symbols represent the emotional state. The size of each symbol represents the amount of activation in each cell. This figure shows the emotional state after state A has increased. Comparing the left and right figures, it can be seen that the occupied area by state A increased, while the areas of the other states became smaller than before. Calculating the number of the occupied area of each symbol, the network gives an output as a four-dimensional value.

Fig. 10. Case study II: Visitor Robot performs at *Arti Visive 2*.

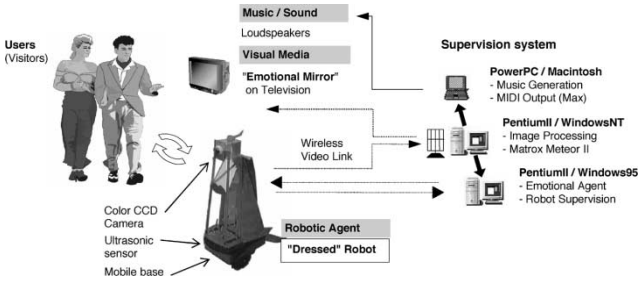


Fig. 11. The supervisor system at *Arti Visive 2*.

E. Performance Demonstration

The developed system has been demonstrated in the interactive art installation at *Arti Visive 2* (Visual Arts 2), a theatrical exhibition held in Palazzo Ducale, Genova, Italy, in October 1998. Sensors allow the robot not only to avoid collisions with people surrounding it, but also to observe the artworks and the visitors in order to interact with them. The robot also has been "dressed" with a scenography for the art installation. On the top of the dressed robot, a small camera has been installed. Fig. 10 shows the working robot at the exhibition site. Fig. 11 illustrates the supervisor system.

As for the music generation, the main four patterns of music that correspond to each emotional state are prepared by a composer. Each music pattern is composed so that the impression would reflect each emotional state. These patterns appear in parallel or simultaneously, not in series. The created music, therefore, is a mixed music of four patterns with particular sound effects.

Visitors can communicate with the robot in several ways. For example, they can approach it, act in front of its "eyes," follow it, ignore it, or become an obstacle in its path. The robot interprets some stimuli as positive, others as negative, causing the evolution of its emotional state. As described above, the emotional state is exhibited by means of the robot's movement, music, sound, and visual media (see also [40]).

V. Case Study III: the iDance Platform

A semiautonomous human cooperative robot that is categorized in the right-upper quadrant in Fig. 1 will be described in this section. Through direct physical contact by a human, the robot follows the applied force by a human performer, displays its own motion, and creates sound and music according to the given inputs from surrounding environment.

Fig. 12 illustrates the developed robot that allows the performer to step upon and to make an action on it. The robot moves in an omnidirectional manner along with a weight shift of the performer. The corresponding sound is fed to the onboard stereo loudspeakers. The input data from the four strain-gage sensors, and the output of the sound and image analyzer modules are given to the music generator and the behavior coordinator. Note that communication data are exchanged through the MIDI channel between the main controller and others.

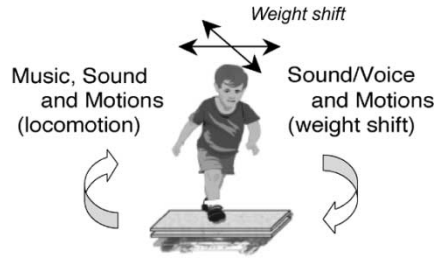


Fig. 12. Case study III: the iDance platform, a semiautonomous robot for music-based human–machine interaction [46].

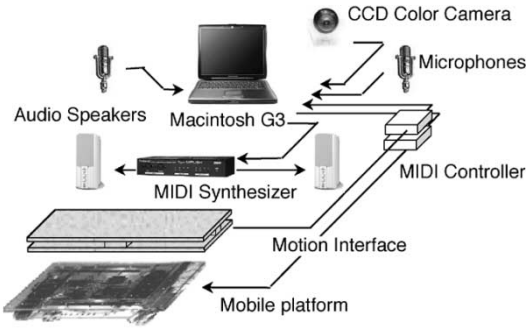


Fig. 13. System overview of the iDance platform. To make the interaction with a human, the system integrates three kinds of communication channels: acoustic (music and sound), movement (the behaviors of the robot), and visual.

A. System Overview

An overview of the developed system is shown in Fig. 13. The system consists of four components: mobile robot, motion interface, main controller, and output devices. In this study, an omnidirectional mobile platform [47] is used for a mobile base. In addition, a motion interface, called a plate, is installed in order to receive external force information. The interface plate enables simple locomotion by a weight shift and force application [48]. Also, a CCD camera and two microphones are installed in order to get environmental visual and auditory information. All these instruments and others, including a Macintosh G3 computer and audio speakers, are installed to make the mobile robot semiautonomous. A number of useful modules for motion devices have added to the Max/MSP architecture. The modules communicate with the robot and motion interface through the MIDI controller to exchange serial and MIDI data.

B. Sensing the External Environment

To have interactions with a human, the system integrates three kinds of communication channels: acoustic (music and sound), movement (the behaviors of robot), and visual. Three input modules will be described in this section.

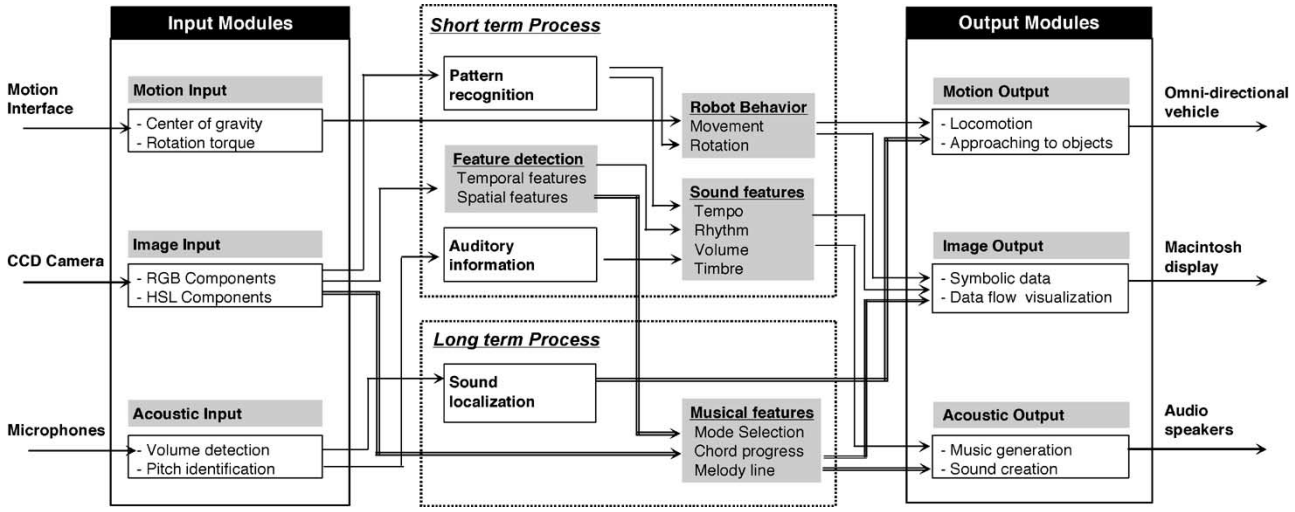


Fig. 14. Dataflow of the iDance platform. Based on the prototype of generated music, sound and musical features can be modified according to the input from sound, image, and motion modules.

1) Motion Interface Module: The first module is an action receiver that has the role of gathering data from the motion interface at the MIDI rate (31.25 kb/s). The interface can obtain the center of gravity data of objects on the mobile robot, which is measured by four strain-gage sensors bonded under the plate. If a user provides a force to the object on the plate, the four strain-gage sensors bonded under the plate measure the center of gravity. Since the strain-gage element changes its resistance value according to the applied load, the applied load can be measured as a change in voltage. The center of gravity is then calculated using these voltage values by a onboard single chip computer. The module on the Max/MSP module receives a data list about the center of gravity and calculates the direction of the movement. In particular, a sudden change of the center of gravity is regarded as a sign of scene change, for example, jumping upon the plate or twisting his/her body on the plate.

2) Sound Input Module: The sound input module is a collection of components for the sound input. The auditory information from two microphones installed on both sides of the mobile platform can be obtained, as well as the *sound analyzer* as described in Section III-B2. In this case study, the sound input is not used much for the music creation and the behavior coordination because the emitted sound and behavior are tightly coupled with the applied force to the robot.

3) Camera-Based Sensor System: The third module is a camera-based sensor system to obtain environmental information and human gestures. This submodule is a modified *image analyzer* as described in Section III-B1. In this case study, however, the obtained image is not used much, just for the robot's rotation so as to track the moving object near the robot. The system allows a human to communicate with the robot with the aid of a small symbolic source such as an LED light or color flags. From this, users can provide a sign to the robot with the location of detected symbol.

C. Robot Reaction

The output module consists of two parts: sound and music generation, and control movement of the robot. Each output module operates under the influence of the input modules. The output parameters are calculated from the input parameters of the external environmental information through two kinds of process component: long-term and short-term reactive ones. The details of the music creation are described as follows and the dataflow is shown in Fig. 14.

1) Omnidirectional Mobile Robot Control: The first output module is the part that controls the robot. At present, two types of controls, active and passive reaction, are prepared.

The former one works as a sort of robot behavior to chase humans. For example, with the aid of symbolic flags, a robot can detect a human and simply follow the symbolic flag. This chasing reaction also is caused by the input from the sound detection module. By using sound localization data, the robot can turn and change the direction by itself. The latter is a kind of tool that a human can use to control the robot. As a passive reaction, on the other hand, when a human pushes the robot, it moves itself along the direction that the force is given. In other words, this module allows a human to show his intention with his action. The communication data format is the same as the MIDI configuration, and sent to the external MIDI controller (motion MIDI). The special hardware translates the MIDI format into the control of the mobile robot.

2) Music and Sound Creation: The second output module is a part of the music and sound generation. We have developed a specific Max/MSP module for the robot. The melody note corresponds to the distance between the center of the motion interface and center of gravity of an object (performer) on it, while the change of the position of the center of gravity corresponds to the MIDI velocity. The kind of instrument changes in a random manner, which is caused by a scene change.

Moreover, some basic modes of the music generation are prepared. Based on the prototype of generated music, sound and musical features can be modified according to the input from the sound, image, and motion modules. The mobile robot becomes active when it is put into the environment where the robot and humans perform. All of the output can

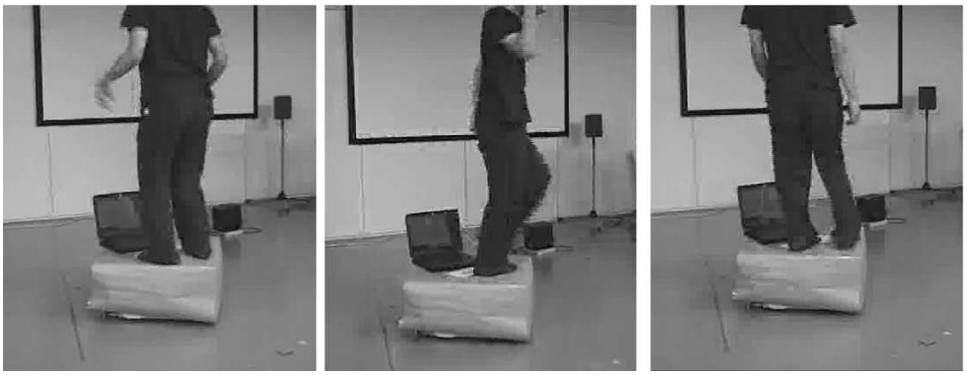


Fig. 15. Performance of the iDance platform: an example of dance performance with the iDance platform. The performer freely makes action on the platform. The mobile base moves according to his weight shift while producing the sound.

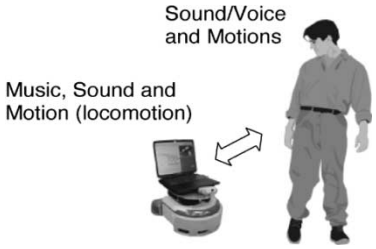


Fig. 16. Case study IV: MIDItro, an autonomous robot for music-based human–machine interaction [49].

then be continuously changed and can also be modified according to both acoustic and visual features of the environment.

D. Performance Demonstration

The performer freely produces action on the platform, which continues to create sound and music according to a given stimuli. As illustrated in Fig. 15, the demonstration showed us that the system performed an interesting interaction between users and robot by using multimodal communication channels (see also [40]). By providing the user's intention to the robot with his/her action, the robot not only reflects the intention with music generation but also with motion. The small motions of the human may be amplified by the robot to make the performance more dynamic and exciting.

VI. Case Study IV: MIDItro

An autonomous robot will be described in this section. The robot autonomously makes an action and creates sound and music according to the obtained sounds and images, and an applied force in physical contact. The performer, thus, moves freely near the robot, interacts with the robot in different manners as illustrated in Fig. 16. The robot is, thus, driven according to the given stimuli from the performer such as a voice, a handclap, gesture, and physical direct contact. This platform is an extended model of the *iDance*. *MIDItro* not only responds to the given input but also behaves in an autonomous manner with regard to its movement and the music creation.

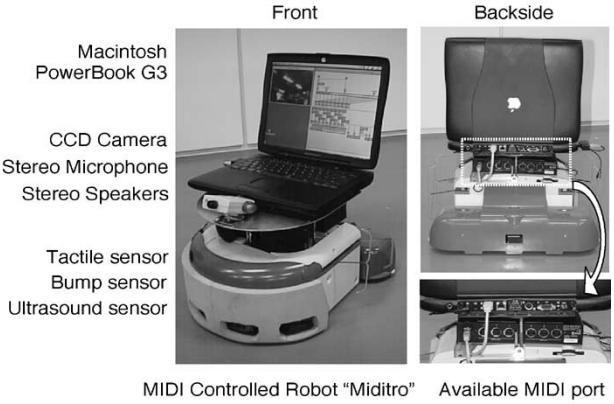


Fig. 17. Case study IV: MIDItro system overview. All the equipment is installed on an omnidirectional mobile base.

A. System Overview

An overview of the developed system installed in a two-wheeled mobile robot is shown in the left figure of Fig. 17. The system consists of four components: a mobile base, main controller, and input and output devices. In this case study, an omnidirectional mobile robot is used for the mobile base. A color CCD camera and two microphones are installed in order to obtain environmental visual and auditory information. All these instruments and others, including the Macintosh G3 computer and audio speakers, are installed to make the mobile robot autonomous. A hardware connector is constructed so as to adapt the mobile base on Max/MSP architecture. The modules communicate with the robot and main controller through the MIDI controller.

The robot consists of two components: an upper installed part and a lower mobile part. The CCD cameras and two microphones for stereo input are attached in front of the upper part on the robot. This robot has bumper sensors on the front and the back, tactile sensors on the sides, ultrasonic sensors on the front, and an encoder, a gyro sensor in the lower mobile part. Therefore, the location of the robot can be roughly calculated using these sensing data. The cushion on the bumper reduces the shock from any collision. A single-chip computer is embedded in the robot in order to handle the low-level control, drive motors, and wheels. For

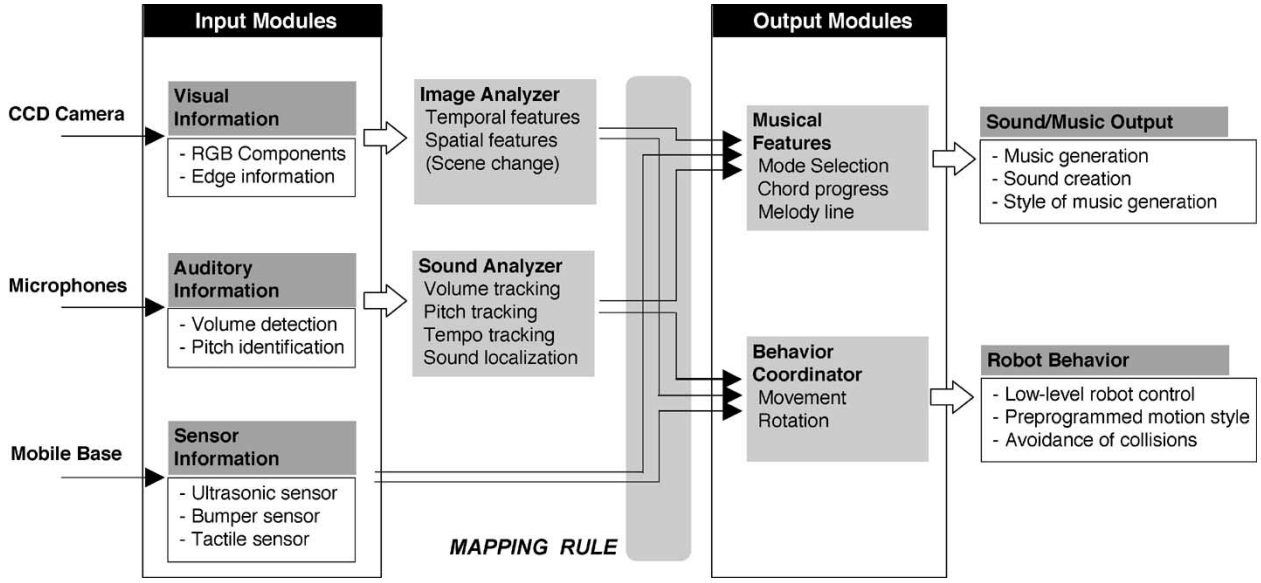


Fig. 18. Processing flow of MIDItro. The mapping rule between input and output modules is determined by composers and choreographers. The output module consists of two parts for the sound and music generation and the control of the robot movement. Each output module operates under the influence of environmental data from the input modules.

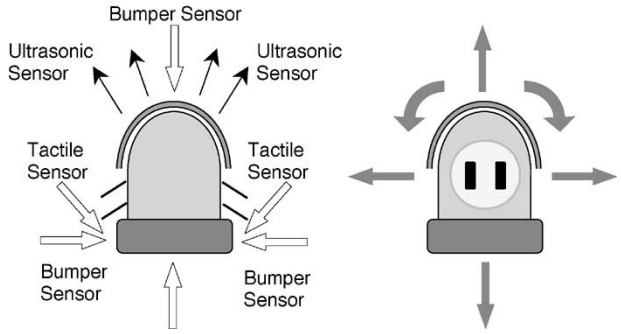


Fig. 19. Sensors equipped with MIDItro. There are three kinds of sensing devices: bumper sensor (binary signal), ultrasonic sensor for the distance measurement, and tactile sensor for sensing data upon physical contact.

the whole system, the size of the robot is approximately 40 cm (width) $\times$ 45 cm (length) $\times$ 20 cm (height).

B. Sensing the External Environment

In order to modify the musical parameters such as melody, backing, tempo, and pitch, three types of information as described below are available to obtain the sensing parameters. The processing flow is illustrated in Fig. 18.

1) Sensor Information: The robot has three kinds of different sensors. The bumper sensor installed on front and back can sense obstacle contact in seven different directions. The four-wire tactile sensors installed on both sides of the robot can sense an applied force. The ultrasonic sensors can measure the distance to obstacles in front of the robot in four directions in the range of 50 to 500 mm. The sensor configuration is illustrated in Fig. 19. These three sensors are used not only to directly link musical parameters but also to avoid obstacles without transmitting MIDI data. For safety, when a bumper sensor senses obstacles, the robot stops and retraces its steps. The other two sensors can be useful for avoiding collisions in advance. These sensing data are converted by the MIDI controller and transmitted to the main controller when the events occur.

2) Visual and Auditory Information: A camera-based sensor system is utilized to obtain environmental visual information and human gestures. By using a modified *image analyzer* as described in Section III-B1, moving image data from the CCD camera are computed to get the spatial and temporal features.

The other module is a collection of components for the sound input. This module can obtain auditory information with two microphones installed on both sides of the mobile platform. The following two specific submodules for this case study were developed for the Max/MSP module. From this, the system allows a user to interact with the robot by using his voice and a handclap. The volume and pitch data of the obtained sound from the installed microphone are calculated, as well as the *sound analyzer* as described in Section III-B2. For example, when the system obtains a human's voice and is able to capture the pitch information, the backing scale will be changed to allow a user to control the high/low chord with his voice.

a) Sound Localization: Simple sound localization has been also realized in one submodule. By using the difference in amplitudes from each microphone, the system can roughly estimate the location of the sound sources. It is not easy to detect exact sound sources with only two microphones. However, because the robot can turn toward the measured target, it helps to capture the exact sound sources more precisely.

b) Tempo Tracking: While users clap hands, another submodule calculates the tempo by extracting the peak of the volume data. The system can synchronize the generated music with the estimated tempo every 20 ms. The player, however, usually does not sustain a constant tempo. The

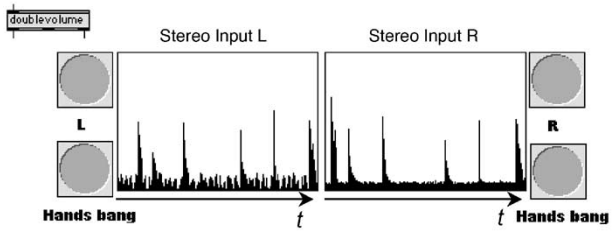


Fig. 20. Tempo tracking object on Max/MSP. While users clap hands, this submodule calculates the tempo by extracting the peak of the volume data.

Table 1
Predetermined Set of Primary Factors for Robot Movement

Robot movement	MIDI Code	Primary Factor
Basic movement		Bumper Sensor
Forward	B-01	Contact at backside
Backward	B-02	Contact at front
Sliding		Tactile Sensor
left direction	B-03	Contact at right part
right direction	B-04	Contact at left part
Rotate		Sound Localization
clockwise	B-05	Source located at right
counterclockwise	B-06	Source located at left
Zigzag motion	B-07	Randomly occurred
Circle motion	B-08	Scene Change

module, thus, renews the next tempo in order to take into account the tempo fluctuation by an experimental threshold. Fig. 20 shows an example of tempo tracking with two microphones. Since the sound is captured every 10 ms, the time difference between each microphone is not found in this experiment. In Fig. 20, the left image represents the sound input from a microphone equipped with the left side, while the right image represents the sound input from the right one. The x axis of each graph represents time t, while the y axis represents the volume of the input sound source.

C. Robot Reaction

1) Mobile Base Control: The robot can move in an omnidirectional manner by rotating the lower mobile part against the upper installed part about the center axis of the robot and by independently controlling two driving wheels. In this case study, the robot is limited to six types of movements such as forward/backward, left/right sliding, and rotate clockwise/counterclockwise. The maximum speed is set to approximately 30 cm/s. A mechanism of collision avoidance is programmed by utilizing the ultrasonic sensors and the tactile sensors.

For the experiment, the primary factors that cause each style of robot motion are predetermined, as described in Table 1. The MIDI code shows the defined addresses for transmission to the computer.

We also demonstrated a direct robot control by the MIDI keyboard. The predetermined robot motions are associated with the notes of the keyboard, as illustrated in the middle of Fig. 21. In this proposed platform, any additional musical instruments that can transmit MIDI data are possible instruments to be associated with.

2) Mapping From Given Inputs to Music: The music is created so as to reflect the given inputs from the external environmental information through two kinds of process components: long-term deliberate and short-term reactive processes. In part of the music and sound generation in a long-term deliberate manner, some basic modes of music composition are prepared as well as *music generator*. Based on these prototypes of generated music, music and sound features such as timbre, pitch, volume, tempo, and style of music generation can be modified according to the input from the sound, image, and motion modules, while a short-term reactive sound is created in a simple manner by outputs from *image analyzer*.

D. Performance Demonstration

Fig. 21 illustrates a performance of MIDItro. Through a number of demonstrations with dancers, the synchronization between humans and the robot can be found (see also [40]-IV). Dancers did not significantly care about the compositional structure of music, but they did care about the variety of the composed music. These demonstrations showed us that the developed multimodal communication channels allow them to make a sophisticated interaction with the robot.

VII. Discussion

Four robotic interfaces in the virtual musical environment are successfully employed in art installations and demonstrations at public exhibitions (for example, *Arti Visive* in 1997, ICMC demonstrations in 1999 and 2000, and robot exhibitions in 1999, 2000, and 2001). These systems can exhibit a "human–robot dance collaboration" where the robot moves in concert with the human performer by sensing the visual and audio information. Each system works as a sort of reflector to create an acoustic and visual space around the moving instrument. The robot can display the reactive motion according to the context of performance to create the human–robot collaborative performance on a stage. Although it is difficult to evaluate and assess the effects of these robots, we believe that the development of these robots that can perform in the real world is a worthwhile subject.

The flexibility of the instrumentation must be considered for the performance systems. The conventional interactive systems have paid less attention to designing the environment for users. The system designer can easily associate the unrestricted relationship between different inputs and outputs because all the components communicate with each other in the Max/MSP programming environment. The proposed platform, thus, provides a useful design environment for artists such as musicians, composers, and choreographers not only to create music but also to coordinate the media including the types of behavior of the robot with the aid of the MIDI network.

As shown by the diagrams of the experimental systems in Fig. 22 (the iDance platform) and Fig. 23 (MIDItro), the

Fig. 21. Performance of MIDItro. The left image illustrates an example of interaction based on the surrounding sounds. In the middle image, the robot is controlled by a MIDI keyboard with the aid of the MIDI network. The right image also illustrates a vision-based interaction.

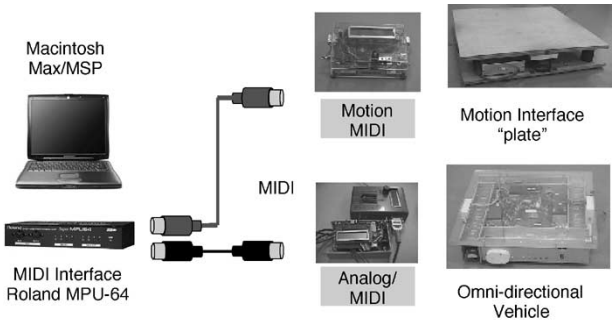


Fig. 22. MIDI network in the iDance platform. Motion MIDI and analog/MIDI are small logic components whose processor is a *Microchip PIC* controller.

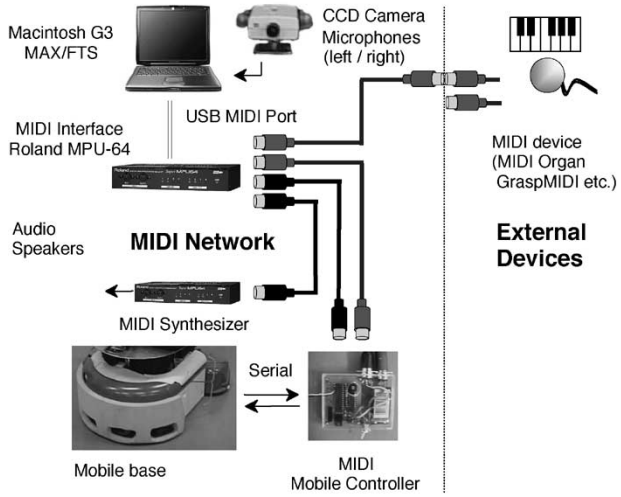


Fig. 23. MIDI network in MIDItro. The MIDI mobile controller is a small logic component whose processor is a microchip PIC controller.

robot can receive data as a control command and transmit data from the sensors to the controller. The microchip converter operates to exchange between these data and the MIDI data. The main controller can control the mobile robot with MIDI data just like musical instruments with a MIDI software environment. By taking advantage of the MIDI format, other MIDI devices can be adopted with the system. For example, we demonstrated a performance with a MIDI keyboard that enables the user not only to play music but also to control the mobile robot.

In the multimodal interaction system for the felicitous performance, the appropriate responses must be required in practically real time. Fast and stable MIDI data transmission is an anxious matter for musicians and scientists. With regard to the robotic applications, the MIDI transmission rate, 31.25 kb/s, is not suitable to ensure a precise control of motors, but is fast enough to exchange data for a behavioral control. The MIDI-controlled robot is still challenging, but recently a commercial production [50] and an artistic work [51] are published. The substantial presence of a robotic interface is one of the possible solutions to make a reaction according to expressive motion of dancers and performers. MIDI-based robotic applications allow them to benefit from the flexibility and scalability of the MIDI devices.

VIII. Conclusion and Future Work

In this paper, four mobile robot platforms to create multimodal artistic environments for music and dance performance have been introduced. The proposed approach to equip musical instruments with an autonomous mobile ability will provide a new type of computer music performance. The developed system reflects environmental visual and auditory information around the human and the robot for the creative and dynamic performance. Since users can provide their intention to the robot by actions, a new style of possible music generation can be provided. We figured out that the system has the capability of creation in the virtual world to extend robot control in the real world.

A human cooperative robot as a partner that makes cooperative work with people will appear in the near future. These robots are required to have a multimodal interface such as visual, audio, and other sensory abilities, to enable them to share information space with humans. It is very important for such robots to have abilities not only to actively work in the human environment, but also to have a flexible and safe interface that needs no specific training or tuning.

In this study, reactive responses in human–machine communications have mainly been addressed. Recently, it is considered to add agents capable of reflecting the users' preferences. The system can be driven not only by the user's intention to the robot but also by the generated music and the sound by itself. When the system starts to interact with a human, the user and system can make a sort of cross perfor-

mance as if the initiative of the performance is continuously transferred between the performer and the system.

We consider that the multimodal musical environment is an interactive space where the robot would behave in response to the given stimuli and its internal state, and the users or performers can continuously interact with the robot. The system, therefore, works as an emotion activator stimulating human creativity. This means that it not only behaves like a human based on the emotional understanding from human movement, but also acts to "activate" humans by integrated outputs. The conventional graphical user interface is not sufficient for interactions in such an environment. It is considered that a substantial interface such as robot and virtual reality is absolutely imperative. We believe that the embodied interaction between a human and the robot will open the next stage of human–machine collaborative musical performance.

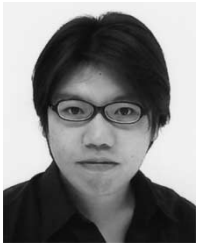
ACKNOWLEDGMENT

The authors would like to thank A. Camurri of the University of Genoa, Genoa, Italy, for his invaluable contribution to this work. The authors would also like to thank T. Hashimoto and E. Nomura, two performers who showed their remarkable talents during the demonstrations; R. Dapelo for composing sensuous music for the *Arti Visive 2* exhibition; and E. Pischedda for developing the scenography intervention ("dress") on the robot in the *Arti Visive 2*. The authors would also like to thank a number of organizations for their partial support of our robot demonstrations. The authors would also like to acknowledge the anonymous reviewers for their helpful comments and suggestions.

REFERENCES

[1] A. Mehrabian, *Nonverbal Communication*. Chicago, IL: Aldine-Atherton, 1972.
[2] S. Hashimoto, "Humanoid robot for Kansei communication," in *Proc. 2nd Int. Symp. Humanoid Robots*, 1999, pp. 156–160.
[3] J. Pressing, "Cybernetic issues in interactive performance systems," *Comput. Music J.*, vol. 14, no. 1, pp. 12–25, 1990.
[4] C. Bahn, T. Hahn, and D. Trueman, "Physicality and feedback: a focus on the body in the performance of electronic music," in *Proc. Int. Computer Music Conf.*, 2001, pp. 44–51.
[5] F. Michaud, A. Clavet, G. Lachiver, and M. Lucas, "Designing toy robots to help autistic children—An open design project for electrical and computer engineering education," presented at the Conf. Amer. Soc. Engineering Education, St. Louis, WA, 2000.
[6] K. Dautenhahn and I. Werry. (2001, Jan.) The AURORA project: Using mobile robots in autism therapy. *Learn. Technol.* [Online]. Available: http://lttf.ieee.org/learn_tech/issues/january2001/#12
[7] L. Hiller and L. M. Isaacson, *Experimental Music*. New York: McGraw-Hill, 1959.
[8] M. V. Mathews and F. R. Mooer, "GROOVE—a program to compose, store, and edit functions of time," *Commun. ACM*, vol. 13, no. 12, pp. 715–721, 1970.
[9] I. Xenakis, *Formalized Music: Thought and Mathematics in Composition*. Bloomington: Indiana Univ. Press, 1972.
[10] Interactive systems and instrument design in music, J. Chadabe and M. Wanderley. (2000). [Online]. Available: http://www.igmusic.org/
[11] J. Paradiso, "Electronic music: new ways to play," *IEEE Spectr.*, vol. 34, pp. 18–30, Dec. 1997.
[12] T. Machover and J. Chung, "Hyperinstruments: musically intelligent and interactive performance and creativity systems," in *Proc. Int. Computer Music Conf.*, 1989, pp. 186–190.
[13] J. Chung, N. Gershenfeld, and M. A. Norris, "A development environment for string hyperinstrument," in *Proc. Int. Computer Music Conf.*, 1991, pp. 150–152.
[14] A. Vidolin, "Musical interpretation and signal processing," in *Musical Signal Processing*, C. Roads, S. T. Pope, A. Piccialli, and G. De Poli, Eds. Lisse, The Netherlands: Swets & Zeitlinger, 1997.
[15] P. Cook and C. Leider, "SqueezeVox: a new controller for vocal synthesis models," in *Proc. Int. Computer Music Conf.*, 2000, pp. 519–522.
[16] P. Cook, "Principles for designing computer music controllers," presented at the NIME Workshop—Proc. ACM CHI 2001, Seattle, WA, 2001.
[17] Y. Nagashima and T. Tono, "'It's SHO time'—an interactive environment for SHO (Sheng) performance," in *Proc. Int. Computer Music Conf.*, 1999, pp. 32–35.
[18] H. Morita, S. Hashimoto, and S. Ohteru, "A computer music system that follows a human conductor," *IEEE Computer*, vol. 24, pp. 44–53, July 1991.
[19] W. Siegel and J. Jacobsen, "Composing for the digital dance interface," in *Proc. Int. Computer Music Conf.*, 1999, pp. 276–277.
[20] M. Wanderley and M. Baffier, *Trends in Gestural Control of Music*. Paris, France: IRCAM, 2000.
[21] J. Paradiso, "The brain opera technology: new instruments and gestural sensors for musical interaction and performance," *J. New Music Res.*, vol. 28, no. 2, pp. 130–149, 1999.
[22] K. C. Ng, "Sensing and mapping for interactive performers," *Organized Sound*, vol. 7, pp. 191–200, 2002.
[23] J. Nakamura, T. Kaku, K. Hyun, T. Noma, and S. Yoshida, "Automatic background music generation based on actors' mood and motions," *J. Visualizat. Comput. Animat.*, vol. 5, pp. 247–264, 1994.
[24] A. Camurri, S. Hashimoto, M. Ricchetti, A. Ricci, K. Suzuki, R. Trocca, and G. Volpe, "EyesWeb—toward gesture and affect recognition in dance/music interactive systems," *Comput. Music J.*, vol. 24, no. 1, pp. 57–69, 2000.
[25] M. Lyons and N. Tetsutani, "Facing the music: a facial action controlled musical interface," in *Proc. ACM CHI—NIME Workshop*, 2001, pp. 309–310.
[26] H. Sawada and S. Hashimoto, "A haptic device driven by grasping force for hand gesture tele-communication," in *Proc. ASME Dynamic Systems and Control Divison*, 1999, pp. 437–444.
[27] S. Fels and K. Mase, "Iamascope: a graphical musical instrument," *Comput. Graph.*, vol. 23, no. 2, pp. 277–286, Apr. 1999.
[28] Lecture for 'Info art', D. Rokeby. (1995). [Online]. Available: http://www.interlog.com/drokeby/install.html
[29] Virtual cage, C. Möller. (1997). [Online]. Available: http://www.canon.co.jp/cast/artlab/pros2/
[30] B. Johannes, "The Intelligent Stage," *Perform. Res.*, vol. 6, no. 2, pp. 116–122, 2001.
[31] C. Breazeal and B. Scassellati, "Infant-like social interactions between a robot and a human care-giver," *Adaptive Behav.*, vol. 8, no. 1, pp. 49–74, 2000.
[32] T. Nakata, T. Sato, and T. Mori, "Expression of emotion and intention by robot body movement," in *Proc. Int. Autonomous Systems*, vol. 5, 1998, pp. 352–359.
[33] SoundCreatures, K. Eto. (1998). [Online]. Available: http://www.canon.co.jp/cast/artlab/scweb/
[34] K. C. Wasserman, M. Blanchard, U. Bernardet, J. Manzolli, and P. F. M. J. Verschure, "Roboser: an autonomous interactive composition system," in *Proc. Int. Computer Music Conf.*, 2000, pp. 531–534.
[35] D. Trueman and P. Cook, "Bossa: The deconstruced violin reconstructed," in *Proc. Int. Comput. Music Conf.*, 1999, pp. 232–239.
[36] S. Takahashi, K. Suzuki, H. Sawada, and S. Hashimoto, "Music creation from moving image and environmental sound," in *Proc. Int. Comput. Music Conf.*, 1999, pp. 240–243.
[37] Y. Gong, C. Hook-Chuan, and G. Xiaoyi, "Image indexing and retrieval based on color histograms," in *Multimedia Modeling*, T. S. Chua, Ed, Singapore: World Scientific, 1995, pp. 115–126.
[38] D. Cope, "An expert system for computer-assisted composition," *Comput. Music J.*, vol. 11, no. 4, pp. 30–46, 1987.
[39] R. Rowe, *Interactive Music Systems—Machine Listening and Composing*. Cambridge, MA: MIT Press, 1993.
[40] Online demonstration—Moving instruments, K. Suzuki. (2003). [Online]. Available: http://www.phys.waseda.ac.jp/shalab/\ ~kenji/demos.html

[41] K. Suzuki, A. Camurri, P. Ferrentino, and S. Hashimoto, "Intelligent agent system for human–robot interaction through artificial emotion," in *Proc. IEEE Int. Conf. Systems, Man, and Cybernetics*, 1998, pp. 1055–1060.
[42] T. Kohonen, *Self-Organizing Maps*. Berlin, Germany: Springer-Verlag, 1994.
[43] A. Camurri and A. Coglio, "An architecture for emotional agents," *IEEE Multimedia*, vol. 5, pp. 24–33, Oct.–Dec. 1998.
[44] K. Konolige and K. Myers, *The Saphira Architecture for Autonomous Mobile Robots*. Cambridge, MA: MIT Press, 1996.
[45] T. Darrell, G. Gordon, J. Woodfill, and H. Baker, "A magic morphin' mirror," in *SIGGRAPH '97 Visual Proc*. New York: ACM, 1997. [CD-ROM].
[46] K. Suzuki, T. Ohashi, and S. Hashimoto, "Interactive multimodal mobile robot for musical performance," in *Proc. Int. Computer Music Conf.*, 1999, pp. 407–410.
[47] S. Hirose and S. Amano, "The VUTON: high payload high efficiency holonomic omni-directional vehicle," in *Proc. Int. Symp. Robotics Research*, 1993, pp. 253–260.
[48] J. Yokono and S. Hashimoto, "Motion interface for omni-directional vehicle," in *Proc. 7th Int. Workshop Robot and Human Communication*, 1998, pp. 436–441.
[49] K. Suzuki, K. Tabe, and S. Hashimoto, "A mobile robot platform for music and dance performance," in *Proc. Int. Computer Music Conf.*, Berlin, 2000, pp. 539–542.
[50] MIDI robot (2002). [Online]. Available: http://www.vutag.com/
[51] S. Jordá, "Afasia: the ultimate homeric one-man-multimedia-band," in *Proc. Int. Conf. New Instruments for Musical Expression*, 2002, pp. 132–137.



Kenji Suzuki (Member, IEEE) received the B.S., M.S. and Dr.Eng. degrees in applied physics from Waseda University, Tokyo, Japan, in 1997, 2000 and 2003, respectively.

From 1997 to 1998, he was a Visiting Researcher at the Laboratory of Musical Information, University of Genoa, Italy, where he was engaged in a project on human–robot communication. From 2000 to 2002, he was a Research Fellow of the Japan Society for Promotion of Science. Since 2000, he has been a Researcher with the Humanoid Robotics Institute, Waseda University. He is also a Research Associate with the Department of Applied Physics, Waseda University. His research interests include neural computing, Kansei information processing, affective computation, computer music systems, and multimodal human–machine communication.

Shuji Hashimoto (Member, IEEE) received the B.S., M.S. and Dr.Eng. degrees in applied physics from Waseda University, Tokyo, Japan, in 1970, 1973 and 1977, respectively.

From 1979 to 1991, he was with the Faculty of Science, Toho University, Funabashi, Japan. He is currently a Professor in the Department of Applied Physics, School of Science and Engineering, Waseda University. Since 2000, he has also been a Director of the Humanoid Robotics Institute, Waseda University. He is the author of over 200 technical publications, proceedings, editorials, and books. He is a reviewer for many journals and conferences on computer music, computer vision, robotics, and neural computing. His research interests are in human communication and Kansei information processing including image processing, music systems, neural computing and humanoid robotics.

Prof. Hashimoto was a Vice President of the International Computer Music Association from 1996 to 2001. He has been Chair of the Promotion Committee, IEEE Tokyo Section, since 2003.