Journal of New Music Research, 28 (1999), No. 4, pp. 351–363

0929-8215/99/2804-351$15.00
©Swets & Zeitlinger

Music Content Processing and Multimedia:* Case Studies and Emerging Applications of Intelligent Interactive Systems

Antonio Camurri

Laboratorio di Informatica Musicale, DIST,
University of Genova, Genova, Italy

ABSTRACT

Music content processing is receiving more and more attention from the multimedia community. On the one hand, research developments in computer music during the last decades are conducing toward more sophisticated and novel musical applications; on the other hand, such research is useful for the development of non-musical applications embedding music content processing to enhance effectiveness and usability. Intelligent user interfaces, games and entertainment, edutainment, and rehabilitation are some examples.

This paper focuses on the role of analysis and synthesis of expressive content in multimedia signals, with particular emphasis on music, integrated with human movement and visual media in the framework of intelligent interactive systems.

Research on "affective computing" in the USA and on "KANSEI information processing" (KIP) in Japan are consolidated research areas dealing with artificial emotions and expressivity. We focus on a third way, where research on expressivity and artificial emotions is considered in the framework of European culture. In this direction, special emphasis is given to qualitative analysis of human movement and its relations with the music signal, starting from previous studies on movement by choreographers (Laban's "Theory of Effort"). We aim at a system that is able to distinguish between the different expressive content of two performances of the same dance fragment. Recent concrete results achieved in the last years, including applications in music theatre and in interactive exhibits for edutainment and museums, are presented briefly.

INTRODUCTION

The multimedia community is becoming aware of the importance of sound and music in multimedia applications (e.g., Dannenberg and Camurri, 1994). Music content processing is achieving importance in a growing number of both musical and non-musical applications and novel applications and market niches are emerging.

In this scenario, our paper focuses on the research area of music and interaction (Chadabe, 1996; Rowe, 1993), on expressive content communication in active spaces where integrated human movement (e.g., of a music performer or a dancer) and visual and music languages concur as a whole perceived entity. With regard to movement, special focus is given to the high-level and qualitative analysis of human movement and its relations with the music signal, starting from previous studies on movement such as Laban's "Theory of Effort" (Laban and Lawrence, 1947) and KANSEI

*Sound examples & color images are available in the JNMR Electronic Appendix (EA) which can be found on the WWW at http://www.swets.nl/jnmr/jnmr.html information processing (Camurri, 1997; Hashimoto, 1997). Ideally, we aim at a system able to distinguish between the different expressive content of two performances of the same dance fragment. Other main research issues include real-time analysis and synthesis of expressivity in music and in movement signals (e.g., dance), multimodal interaction, and intelligent interfaces where sound and music as well as visual and gesture languages, and mobile (robotic) scenographies are the components of an integrated multilevel communication. We do not address research issues on music interpretation (Dannenberg and De Poli, 1999) in this paper.

These intriguing research issues allow researchers and artists to approach music tasks such as composition, performance, and the design of novel music and multimedia applications in a different perspective. For example, expressive content parameters can be used for more sophisticated querying in music databases.

A growing number of musical as well as non-musical multimedia applications require intelligent, multimodal user interfaces, i.e., capable of addressing different human senses — therefore including sound and music as a first-class communication channel to convey a higher amount of information. At the same time, a common platform for the representation of sound and music objects on the one hand, and visual and gesture objects on the other, is a particularly interesting research topic addressed in several composition and performance environments.

The modeling, analysis, and synthesis of artificial emotions, expressivity, and more in general KANSEI information (Camurri, 1997; Hashimoto, 1997) are other crucial topics in various classes of multimedia applications involving music content processing. We envisage application areas such as interactive music, theatre, art, and museum installations, possibly including robots on stage, and several other emerging application niches (Camurri, 1995) like interactive edutainment, applications for museums and exhibitions (e.g., involving autonomous robots playing roles similar to an intermediary between visitors and the exhibit to attract, entertain, stimulate, and instruct visitors), interactive entertainment (e.g., an interactive discotheque where dancers can influence and change the music, games like dance-karaoke where "the better you dance the better music you get"), interactive tools for aerobics and gymnastics, rehabilitation tools (e.g., agents supporting therapies for mental disabilities, where it is important to have non-intrusive environments in which patients can establish a creative interaction), tools for teaching by playing and experiencing in simulated environments, and tools for enhancing the communication between new products or ideas in conventions and "information ateliers" (e.g., fashion shows). Furthermore, the analysis and synthesis of expressive data in music and multimedia signals can contribute to the creation of more sophisticated content-driven queries in music and multimedia databases.

We explored and developed research as well as applications in some of the above-mentioned areas (Camurri, 1995; Camurri and Coglio, 1998; Camurri and Ferrentino, 1999; Camurri and Leman, 1997; Camurri et al., in press). Our background includes previous research projects and systems, e.g., MANI (Camurri et al., 1986, 1987) and HARP (Camurri et al., 1994).

In this paper, we introduce some basic research issues and then focus on two of our recent projects: (i) the *EyesWeb* project, whose main goal is to explore extensions of the music language toward gesture and visual languages and (ii) the *Music Atelier* project, which consists of a set of integrated interactive games (agents) for science centers, also based on full-body physical interaction, recently applied in the "Città dei Bambini" (literally, "The City of Children") science center for children.

AGENTS AND ARTIFICIAL EMOTIONS

One intriguing issue concerns the development of cognitive and then computational models able to capture emotional and expressive content, with a particular focus on non-verbal communication (e.g., music and gesture). Research on music interpretation is a well-known field in the computer music community (e.g., Dannenberg and De Poli, 1998). Reports on artificial intelligence exist that discuss cognitive models for agents with emotions (e.g., Sloman, 1998). Reports on "affective computing" (Picard, 1997) and "KANSEI information processing" (Camurri et al., 1999; Hashimoto, 1997) concern research on agents with personality and emotions seen from different cultural perspectives. One agent (Riecken, 1994) is a computer system, possibly equipped with a physical "body" (e.g., a robot), embedding rational (or cognitive), emotional, and reactive capabilities and able to interact with the external world, including other agents.

Camurri and Coglio (1998) proposed a general, flexible, and powerful architecture to build emotional agents, i.e. software agents embedding models of communication by means of the metaphor of artificial emotions. By artificial emotions, we mean information and computations achieving functionalities that relate to conveying emotions to humans, thus allowing more effective, stimulating, and natural interactions between humans and agents. In our approach, we consider the problem of the *expression* of emotional content rather than the (far more ambitious!) problem of synthesizing agents able to feel emotions. An emotional agent integrates different components: input processing from the surrounding world (e.g., motion capture, sound feature recognition), cognitive (or rational) processing, emotional processing, reactive processing, and output processing (generate sound, music, visual media, control actuators). The external world can be a performance stage or a museum exhibit — real, virtual, or a mix of both — populated by artificial as well as human agents that are able to interact, dialogue, cooperate, compete, etc., each contributing to the performance. Camurri and Coglio (1998) described the components of the architecture in detail as well as the data through which they interact with each other and with the external world. The integration of emotional and the other components is of particular interest. The main role of the emotional component in our architecture is to influence the processing of the other components by means of parameters containing information about the emotional state, updated by the emotional component as its state changes. The emotional component consists therefore of a state, a number of processes which intervene at different levels to influence the other components, and a number of processes that influence the emotional component of the other modules. These processes occur at multiple levels. Therefore, the emotional component is tightly integrated with the other components and, in particular, the rational, input, and output components. The model includes processes between components which govern the dynamics of expressive behavior. This also seems coherent with Hashimoto's proposal of the model of KANSEI (1997), where KANSEI is described as a "process" rather than a "state".

The expressive and emotional state, among the many possible, consists of coordinates of a point in some "space". The coordinates can change step by step according to emotional (possibly music, movement) stimuli or even according to some physical metaphor where emotional stimuli constitute forces acting upon a point mass. The emotional space is usually partitioned into zones characterized by synoptic symbolic names (e.g., "brightness", "darkness", "excitement"). Thus, different stimuli tend to move the mass towards different zones. Borders among zones as well as emotional coordinates and forces can be fuzzy. Two instances of this kind of computational model can be found in Camurri and Ferrentino (1999) and Camurri et al. (1997). At present, we are successfully employing the architecture sketched above in various real-world multimedia-multimodal systems.

KANSEI INFORMATION PROCESSING

KANSEI information processing (KIP) is an emerging research area and industry from Japan (Hashimoto, 1997). The concepts behind KIP are strongly tied to Japanese culture. KANSEI is a Japanese word that does not have a direct counterpart in Western languages; in other words, every attempt at translation captures just some of the aspects of KANSEI. The concept of KANSEI is strongly tied to the concept of personality and sensibility. KANSEI is an ability that allows humans to solve problems and process information in a faster and more personal way. In every action performed by a human being, traces of his/her KANSEI can be noticed; this is also true for his/her way of thinking and of solving problems and of his/her personality. Therefore, KANSEI is not a synonym for emotion, although it can be related to emotion. It refers to the human ability to process information in ways not only logical (Hashimoto, 1997). KANSEI is related to both problem-solving tasks and information analysis and synthesis. Through his works or performances, an artist expresses his KANSEI; he leaves traces of his KANSEI in his product, his message. A skilled actor or dancer can mimic the KANSEI required to make the character he is simulating more believable. The person who sees the work of the dancer/actor or the masterpiece of an artist will use his KANSEI in order to evaluate it, to extract meaning and information about the perceived KANSEI. According to Shuji Hashimoto, it is possible to grasp one of the aspects of KIP as a process that permits us to personalize the way in which a message is sent, an action is performed, choosing solutions suitable for the personality and sensibility of the performer. The person who perceives the results of those actions will use his KANSEI ability to extract details about the solutions adopted by the performer, thus obtaining details or hints about the personality and the KANSEI of the performer.

REVISITING LABAN'S THEORY OF EFFORT

The European choreographer Rudolf Laban developed a method to describe movement that he called "Theory of Effort" (Laban and Lawrence, 1947). Effort is a quality of every movement. The effort present in a gesture and its variations during a movement performance convey expressive information. Laban aimed at understanding the factors of the movement that are perceived by the observer and that carry expressive information independently from the specific gesture performed. The expressive content is not only a feature of the kind of gesture performed (a punch, a caress), but it is mainly related to the way it is performed. The variations of some movement factors concur in giving a gesture its expressive power. Laban studied those factors and collectively named four of them "Effort". It is also possible to think about effort as a property of every movement, described by a set of four parameters.

In our research, we built a model of movement description inspired by this part of Laban's work. This might be considered a sort of implementation of the model of KANSEI filtered by European culture. Laban's "Theory of Effort" can be applied at three different levels of complexity. At the basic level, the four parameters describe effort as pertaining to a gesture. They can assume just two values. We refer to this level as "Basic Effort". The four parameters and the values they can assume are time (sustained, quick), space (flexible, direct), weight (light, strong), and flow (fluent, bound). The movements "sustained" in time have a perceived long duration, maintaining this feature in time: e.g., pressing against an heavy object in order to move it is an example of movement with sustained time. "Quick" refers to quick and sudden movements, like the reaction that occurs when touching something that is really hot. The space factors define a movement whose path in space is either direct (follows a straight line: "direct space") or curved (follows a curved line: "flexible space"). "Strong weight" refers to movements performed using strength, as, again, pressing against an heavy object. The opposite is represented by movements during which no apparent strength is used ("light weight"). The following description regarding the flow factor is taken from Laban's book *Effort*: "Movements performed with a high degree of *bound flow* reveal the readiness of the moving person to stop at any moment in order to readjust the effort if it proves to be wrong, or endangers success. In movements done with *fluent flow*, a total lack of control or abandon becomes visible, in which the ability to stop is considered inessential." Laban then defines eight basic efforts (slashing, gliding, pressing, flicking, wringing, dabbing, punching, floating), which are the eight possible combinations of the first three factors (time, space, weight). During a sequence of movements, the effort vector describing it moves in a space called "effort space" and assumes one of the possible states.

The second level of complexity is called "Secondary and Stressed Effort" and takes account of the fact that in performing certain movements one component of the effort may be predominant with respect to the others, which are thus secondary. The basic model is extended and the effort components assume analog values. At the third level, called "Incomplete Effort", Laban takes account of the fact that some of the components of the effort can assume such a secondary role that they almost disappear, they are barely perceived or not perceived at all. In describing effort space, Laban almost neglected the flux parameter, because it is too difficult to directly evaluate, being more complex than the previous three. The weight parameter can also be difficult to evaluate. Usually the level of strength applied in a movement can be appreciated by noticing the modifications that the movements exert on the surrounding environment, for example, measuring how much a weight is lifted or shifted. However, in the case of free body movements, the performer must mimic the use of strength through the use of the appropriate muscles' counter tensions. Weight is therefore not an easy parameter to evaluate directly, specifically using an automatic system.

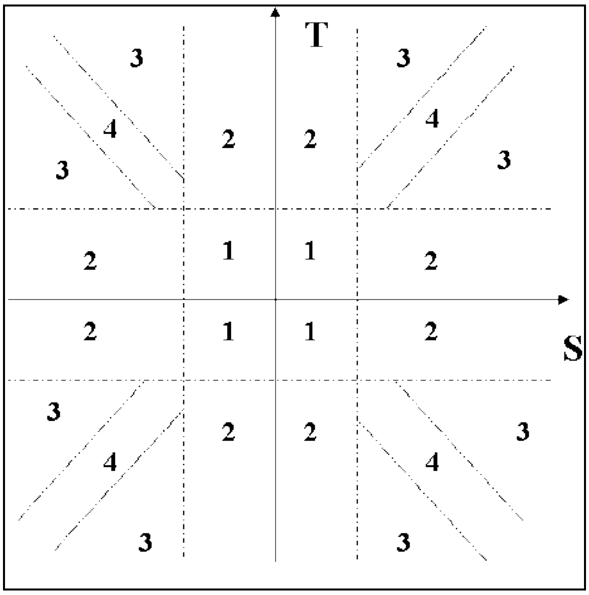

Fig. 1.

An example of a computational model of effort

We developed a computational model of effort space, aiming at a better understanding of the effort theory (Camurri and Trocca, in press).

A representation of our model is depicted in Figure 1. The model takes into account just two effort parameters, space and time. However, it can be extended to include the other two components. Space (S) is intended as a measure of the directness of the path followed in space by the movements, while time (T) is a measure of the impulsive or sustained nature of a movement.

The effort space has been divided into various regions:

(1) regions where the characteristics of effort are not well perceivable;
(2) regions where just one of the characteristics is perceivable, while the others are not (incomplete effort);
(3) regions where one of the components is stressed, while the others are perceived as secondary (stressed and secondary effort); and
(4) regions where all the components are perceived with the same intensity (basic effort).

The type 4 regions (basic effort) are represented as large stripes to take account of a little fuzziness in perception. This model is still very rough because it uses a linear threshold approach, while real perception might work in a more complex way. Furthermore, a better representation of the model can be obtained using polar coordinates. It is, however, useful in order to conduct some basic experiments and to attempt building a method to evaluate human movement that is both quantitative and qualitative.

A *microdance* in our approach refers to a minimal set of movements that permits us to create a context suitable to study the effort and the expressivity of the movement performance. With the help of experts, in particular the choreographer Giovanni Di Cicco, we obtained a set of recorded sequences of movements whose total duration ranges from 20–30 seconds to 1 or 2 minutes that can be, for example, performed conveying different expressive intentions. This means that the same sequence of movements can be performed smoothly or rigidly, lightly or heavily, etc., enabling us to study differences and invariants and, in particular, the different features of effort, starting at time and space dimensions.

The main focus is, therefore, on describing the quality of movement. Variations of effort during the movement performance are of particular interest to understanding movement. Effort can be applied to the analysis of a single part of the body as well as to the movement of the whole body as a point moving on the stage. Laban's *Kinesphere*, also referred to as *personal space*, is the circumference which can be reached by normally extended limbs without changing one's stance, that is, the place of support. The whole space surrounding the kinesphere (i.e., the environment in which the act of movement is taking place) is referred to as *general space*.

THE INTEGRATION OF MUSIC AND GESTURE

Gesture intervenes in different aspects and levels in music performance: the most intuitive concerns the simplest causal relationships typical of musical (hyper)instruments, that is, a local — in space and time — cause-effect process. It can be enriched by opening the observing window in space, in considering the human body action (full-body vs. single joint), and in time (historical and contextual data, preceding and future events and gesture). A single event (the pushing of a switch or an analog sensor) used to produce a note might not be sufficient to capture the deepest expressive content. The movements of preparation before the gesture (e.g., the activation of antagonist muscles in the opposite direction of the intended movement or the anticipation of the event by means of subevents), the past events in the performance (e.g., to evaluate the influence of the preceding gesture on the current event), and the tension in the movement toward the next event are of crucial importance in the design of the interaction. These aspects contribute to a unified and natural perception of gesture and sound.

Expectation and anticipation

The previous aspects can also be seen from the viewpoint of the audio-visual language of films. For example, in the action of a hit of a hammer you can hear its *anticipation* by the artificial amplification of the sound of the wind moved by the hammer while falling to create a stronger tension toward this event on the public. In general, audiovisual language in the cinema can be very interesting to conceive deeper mechanisms to integrate visual/gestural and music languages, such as amplification of anticipation, emphasis, and violation of expectations between audio and visual processes.

It is interesting, although in a different context, to mention the work of Brenda Laurel on the mapping of drama theories from the theater to the design of human-computer interfaces (Laurel, 1993).

A deeper analysis of movement is needed to overcome naive uses of interaction mechanisms which can often be noticed in art installations and hyperinstruments. A structured analysis can help in creating anticipation mechanisms, violation or satisfaction of expectancies, or a contextual response from a gesture by means of dynamic control. As a very simple example, we recently used floor tiles onstage for a music production. Usually, digital or analog floor tiles are used, so, when pressed, you can obtain either an MIDI trigger or a continuous control reflecting the weight pushing on the tile. Our choice was to use two types of sensor in every floor tile: an analog weight sensor (Interlink FSRs) and a proximity sensor (simple electronics analogous to Theremins, i.e., responding to the approach of a part of the body). In this way, the composer receives the information on the velocity, the distance from the tile (by means of proper analysis algorithms), in order to control the event of "hitting the floor tile" with a richer information on its anticipation, including violation of expectation (e.g., a strong speed in approaching followed by a sudden stop).

Interaction metaphors: active spaces as "Characters"

This simple mechanism and these concepts can be generalized and structured situations in active spaces, also with regard to the previously described concepts of KANSEI and effort as a means to deepen the mechanisms and the language of interaction. This is one of the main goals of our EyesWeb system.

Another related important aspect which intervenes in interactive systems concerns the interaction metaphor, which in the case of "virtual active spaces", may need to go beyond the concept and metaphor of musical instrument. In virtual environments (VEs), the most effective human-VE communication often implies the complete re-conception of the VE, the shape and appearance of objects, and manipulating and interacting with them in a completely different way with respect to analogous environments in the real world, even in non-artistic contexts (e.g., the "virtual city", the "virtual shopping mall" on the web).

Analogously, in a virtual environment made of sound and music, the interaction mechanisms and metaphors are not necessarily the same as in the real world, e.g., the musical instruments.

It is important to mention emerging metaphors which move from hyperinstruments to a sort of dialogue between performers/dancers and the active spaces designed as their counterpart. For example, if a dancer changes his "style" of movement, say, by gradually reducing the speed of a harsh movement toward a smoother gesture, the active space/agent will adapt by continuously changing its behavior toward a different *context* (*continuously* and in a time interval proportional to the amount of change in the dancer's style of movement). Transformations can mean a continuous change both in the sensitivity, the interpretation of gesture, and the focus of attention, i.e. a change in the set of movements and gestures observed by the system as well as changes in the causality mechanisms between movement and sound, music output. The dancer has the role of the creator and "moulder" of the "characters"/active spaces emerging on the stage. After the creation of a character, the dancer may interact with it as well as leave it and move toward other places on the stage, e.g., to create other clone-agents.

One of the most important features of our system is the ability to work on a set of *expressive* and *KANSEI* parameters, including different aspects related to the expressive content carried by a movement or a gesture. In doing so, we aim at adding dimensions to our analysis of movement, providing a new paradigm to the composer and a richer interaction with the performer.

The designer of the performance introduces into the system his/her sound and music knowledge, compositional goals, and aspects of integration between music and gesture, including a model of interpretation of gestures and their integration with music parameters, and decides on the amount of (possible) degrees of freedom left to the agent with regard to the generative and compositional choices. This is directed toward extensions of music language with action, gesture languages, and possibly visual and animation languages.

EYESWEB

The EyesWeb project aims at supporting composers, performers, and users in general with models and tools for integrating movement, music, and visual languages, with a particular focus on expressive content. EyesWeb software supports the development of and experimentation with applications of interactive systems. Our main focus is not (only) toward symbol recognition from a posture or gesture vocabulary, but rather toward the extraction of high-level parameters on expressive content in the performance. Existing systems are rather limited from this point of view. Ideally, we want a system that is able to support the design of applications that distinguish between the different expressive content of two performances of the same dance fragment. Our computational model of Laban's "Theory of Effort" described above is available as a set of EyesWeb reusable software modules for movement analysis.

The EyesWeb software, written in MS Visual C/C++ for Win32 platforms, consists of a visual language and a set of libraries of reusable modules which can be used to build patches as in common computer music languages inspired by analog synthesizers.

Basic module libraries include input and movement capture, filters, active modules, observers, and output modules. Movement capture and input modules have been developed for different sensor systems: environmental sensors (e.g., video cameras) and wireless on-body sensors (e.g., accelerometers, FSR, floor tiles). Low-level filters (e.g., preprocessing, denoise, etc.) and high-level filters (e.g., the module to extract the barycentre coordinates of a human figure; the module for evaluating equilibrium) are also available. Active modules have an internal dynamics; that is, they receive inputs like any other module, but their outputs are asynchronous with respect to their inputs. For example, an "emotional resonator" able to react to the perceived expressive content of a dance performance through an internal model of expressivity or artificial emotions may have a delay in activating its outputs due to its actual internal state and processing. Observers (filters or active modules) are modules or patches that are able to extract high-level information, e.g., expressive content, Laban's parameters. Output modules include low-level (e.g., MIDI, network, DMX, etc.) and high-level output modules (e.g., synthesis of expressive content able to modulate a score performed in real-time via Max or SuperCollider).

The problem of how to dynamically adapt a patch, that is, how to control the level and laws of interaction in the active environment, is a crucial issue (Camurri and Ferrentino, 1999; Camurri and Leman, 1997). It usually depends on the evaluation of observer outputs, which can cause proper feedback to activate/de-activate modules and tune parameters.

End users (composers, artists) can directly assemble modules in a patch: they do not need the presence of a programmer, but can simply reutilize the library modules and patches provided by EyesWeb as, for example, those described in the next sections.

New modules can be created in a simple way by C++ programmers by means of "Wizard" software, which guides the expert user step by step in the definition of a new module, to embed a user-defined algorithm into a new module usable in EyesWeb. The Wizard asks the user for the input and output data types of the new modules, makes available to the user a code area to put the user's source code, and automatically generates the COM and surrounding code necessary to transform it into an EyesWeb module. EyesWeb is based on the MS COM/DCOM standard. Details on such standards are hidden by the Wizard from the programmer of a module by generating automatically the necessary COM code.

In Figure 2, the graphic window of an output module is shown. It is usually used as a test module, since it provides a visual feedback from the recognition low-level movement capture and low-level filters. It summarizes visually in a single window the available parameters from a single video-camera (b/w infrared): the occupation rectangles for the human, the barycentre of the two-dimensional silhouette, and the peripheral barycentres (ideally, shoulder, elbow, hand, knee, ankle). Taking this information from a frontal and lateral camera, a three-dimensional approximation can be computed. For each camera, we have 10 vectors with their origin in the main barycentre, whose modules and phases are available as inputs for other modules (filters, observers, outputs). Other EyesWeb filter modules compute relative and absolute velocities and accelerations, kinetic energy of the movement (by means of heuristics approximation of the mass in terms of weighted pixels according to the body shape from the two camera images).

The *equilibrium detector* is a module defined in terms of the displacement of the lower part of the body lying on the floor and the distance between the x-coordinate of the barycentre of the body and the center of the maximum occupation rectangle of the body. Our heuristic algorithm takes into account different cases (e.g., crouched, lying down, etc.). Equilibrium is related (inversely) to the *tendency to immediate movement* parameter: roughly, the more the feet are spread (or the parts of the body on the floor) and the more strongly they are placed on the floor, the more difficult it will be to have sudden and fast full-body movements in the immediate future.

We call *Observer* (O) a high-level module or a patch able to reconstruct a view or understand a particular aspect of the style of movement. The Os delegated to estimate a particular aspect of the expressive content of a dance performance usually study the dynamic properties of movement, e.g., our computational model of effort. To this aim, the following two issues are particularly relevant: (i) the knowledge of the physical laws of human movement (e.g., kinematic constraints of human body, typical acceleration models of motion gestures of parts of the body) allows us to identify the atomic movements components of a movement pattern and (ii) computational models of artificial emotions, which can be used in the evaluation of the expressive content of movement, to build interpretations from the data analyzed by Os. Such modules act as "emotional resonators" on a subset of the input data. The dynamics of the computational models of artificial emotions we adopted depend on past stimuli, future expectations, and the agent internal state. We developed algorithms based on artificial emotion models (Camurri and Ferrentino, 1999) and on Kohonen's SOM (Suzuki et al., 1998). In our approach, the two previous issues are the conceptual link to Laban's theory of movement.

A number of modules and patch libraries are currently available in the system prototype to the users of the system. Figure 3 shows an EyesWeb patch at work in our laboratory at the Opera House Theatre of Genova.

EyesWeb patches can be classified into two categories: observers studying the gestures performed by the dance *inside* the kinesphere, and observers able to study the movement of the kinesphere as a whole entity in the general space.

A *posture analysis observer* is based on back-propagation neural networks (NNs). It simply extracts static information from each frame. Several instances of this agent can be active at the same time, each associated with a different camera and, for each frame, with each different human figure in a single frame extracted from the preprocessing module. The NN can be trained directly by the user (e.g., the choreographer or the composer) with sets of stereotypical postures. The observed human figure in the current frame is evaluated as belonging to one of the learned stereotypical posture or clusters.

Another observer evaluates the *tendency to movement* parameter. It captures information about the steadiness of the position of the dancer. Roughly, the more distant and strongly placed the feet are on the floor, the more difficult it will be to have sudden and fast full-body movements in the immediate future. From the point of view of this O, the “tendency to movement” is a function of the distance between the feet and the symmetry of the body barycentre.

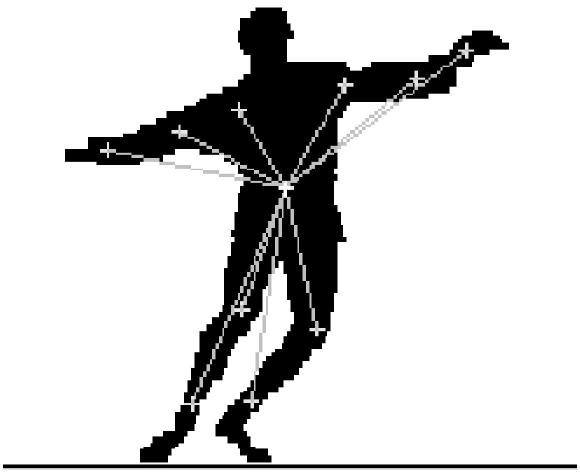

Fig. 2.

Fig. 3.

The *microdances observer* uses color video-cameras. Its goal is to follow parts of the body to study expressive content. This patch concerns the analysis of the same dance fragment (microdance) performed by different dancers or by the same dancer but with different expressive intentions. In this way, we aim at understanding which parameters have to be studied in order to recognize such differences. The current implementation studies the curves of the velocities of each tracked body part and matches them with a model of typical movement velocity curve extracting the parameters necessary to fit the prototypical curve to the observed data. By confronting parameters extracted from curves belonging to the same microdance with different expressive intention and from curves belonging to different microdances performed with the same expressive intention, we tried to build a knowledge database in order to classify a dance performance observed in real time. This approach is in certain aspects analogous to the well-known approach to perceptual analysis in psychology, e.g., in listening and music expression analysis (De Poli et al., 1997).

Other patches are used to evaluate Laban’s effort components space and time as has been shown (see the space-time map in Fig. 1). The patches deal with both movement of a single joint in the kinesphere and the kinesphere in the general space. Another patch evaluates mobility, stability, and lability of movement. An observer module allows

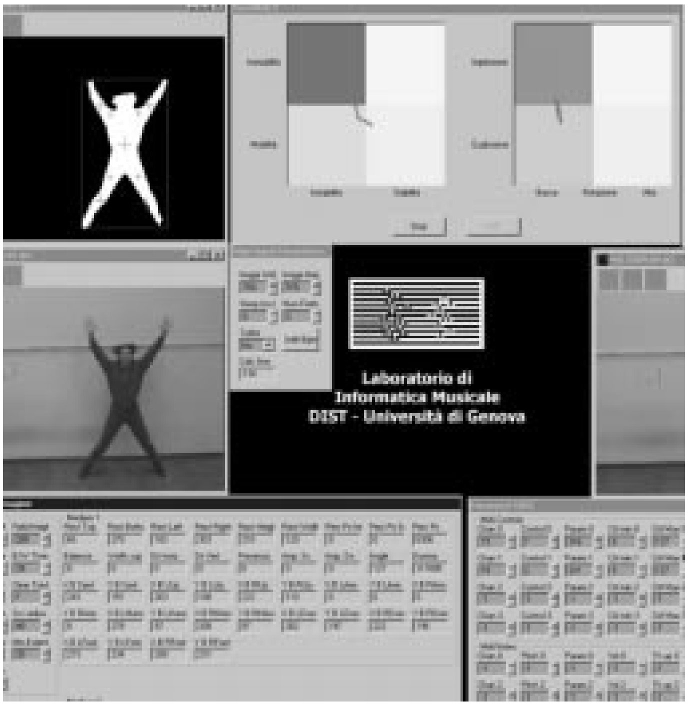

Fig. 4.

the mapping in real-time of a graphical representation of the instantaneous state (or "style") of the movement of the dancer through an animated sequence of points moving on a two-dimensional space (see the EyesWeb patch in Fig. 4).

THE MUSIC ATELIER PROJECT

The Music Atelier project concerns the design of interactive games for museum exhibits, science centers, and children's museums. An example of our project is available in a permanent exhibit at *Città dei Bambini* (literally "children's city"), a science center in Genoa (Italy) where children, by playing interactively with various devices and games, learn concepts in the fields of physics, music, biology, and others. Our laboratory has conceived and realized the Music Atelier of this exhibition. The Atelier at Città dei Bambini consists of five games characterized by multimedia-multimodal interaction involving music, movement, dance, and computer animation. The software for our Atelier is modeled as a population of five communicating "emotional agents", one for each game (Camurri and Coglio, 1998).

In the first game, called *Let Your Body Make the Music*, visitors can create and modify music through their full-body movement in an active sensorized space. Such a biofeedback loop helps visitors learn some basic music concepts (e.g., rhythm, orchestration). Another game, *Cicerone Robot*, is a navigating and speaking (as well as "music performing") robot, that plays the role of a special visitor, explaining and interacting with children (e.g., how to play with other games) and at the same time playing with visitors. The robot can behave by changing its mood and character in response to whatever is happening in the Atelier. For instance, it gets angry if visitors do not play the games the way it had explained them. Its emotional state is reflected in the way it moves (e.g., swaggering, nervous, calm), its voice (e.g., inflection, different ways of explaining things, from short and tense to happy and lively), the music it produces (e.g., happy, serene, noisy), and the environmental lights in the Atelier (e.g., when sad, the light becomes blue and dim; the angrier it gets, the redder the light becomes). The main educational goals of this game are to gradually present artificial intelligence concepts and to rid the idea in children (and adults) of science fiction robots by introducing them to a real robot.

Other games in the Atelier are *Real and Virtual Tambourines* (a set of three percussive musical instruments: only one has a real membrane; the sound source – vibration of the membrane – in the other two is computer-simulated and controlled by sensors), the *Musical String* (explores the nature of musical timbre), and the *Harmony of Movement* (explores some principles of musical language without using traditional notation, but by a visual metaphor with animated characters, a sort of little colored fishes on a computer screen instead of notes on a staff).

As for the hardware and software architecture, the games are implemented as emotional agents. Messages exchanged by agents, besides serving to inform the robot about how visitors are playing the other games, can be useful to dynamically perform mutual adjustments (e.g., adjusting sound volumes of different games to avoid interference) as well as to have the Atelier work as a whole integrated game, where the working of one game can depend on how previous games were played in order to achieve a coherence and continuity in the visitors' tour in the Atelier.

EyesWeb is used to implement interactive games as specific patches. We recently developed an experiment on robot-human communication in a public performance. A mobile robot on wheels (Pioneer 2 by ActivMedia Inc, equipped with further wireless audio and video devices), an evolution of the one used at Città dei Bambini, was placed in an art installation and left free to interact with the public (Arti Visive 2 Museal Exhibition, Palazzo Ducale, Genova). The robot itself was a visitor to the exhibition and embedded a model of artificial emotions, which was influenced by the stimuli it received from the external world. It reacted by generating in real time music, a video (a "digital mirror" software able to mould the video signal acquired by the robot) presenting what the robot saw from its on-board camera mirrored on a large screen and deformed by its emotional state. Another ongoing project deals with the interaction between a robot and a dancer at the event "L'Ala dei Sensi" (Ferrara, November 1999). The goal is to experiment with robot-human communication and, in particular, the synthesis of expressive movement, although simple and constrained by the robot hardware.

PERSPECTIVES ON MUSIC AND MOVEMENT LANGUAGES

This paper raises issues and new perspectives on the integration of music and movement languages. In this scenario, the composer and director of the performance in live electronics (Vidolin, 1997) has to redefine his role and co-operate with the dancers or music performers during the performance. Furthermore, parameters based on the qualitative analysis of movement, on expressive content, and therefore on deeper interaction metaphors, allow the composer and director of the performance to explore new directions of the music language. We can conceive sophisticated examples of interaction in which communication between a dancer and the active space can evolve, based, for example, on learning, imitation, and dialogue: the interaction metaphor can therefore vary from "musical instrument" to "orchestra conduction", to different kinds of dialogue, social interaction. The role of the composer is now to define the "character", the "perception system" of such a character/active space the dancers interact with, and its "learning capabilities". The composer defines in this way musical objects as active entities, deciding which degrees of freedom he leaves to the performer and which kind of interaction and evolution of the music object. It is worth noting that this does not imply necessarily degrees of randomness or stochastic processes in the interaction process. Our guess is that agent architectures and models of artificial emotions and expressivity may contribute to new paradigms in composition and performance. The modeling of composition and performance environments in terms of communities of emotional agents has already been discussed in the literature (e.g., Riecken, 1994). In music composition, our hypothesis – and one of the motivations of our work – is that models of artificial emotions and expressivity might be both a conceptual and a realization platform to investigate new paradigms. For example, theories of the composer Gerard Grisey define a music domain in terms of living entities: his view on sound as *être vivant* rather than *object* seems conceptually close to our view on artificial emotions and emotional agents as living beings during a performance. Our choice of Laban's effort theory is another crucial point in our approach to face these issues.

Applications based on these ideas have been explored in simulations (Camurri et al., 1986, 1987), developed in real-time systems (Camurri, 1995; Camurri and Ferrentino, 1999; Camurri and Leman, 1997) and experimented on by composers and choreographers in public events.

As an example, we recently used EyesWeb in the music theater work "Cronaca del Luogo" of Luciano Berio for the opening of the 1999 Salzburg Festival (July-August 1999). Here, we will briefly address the interaction issues and interface technology developed for Berio's work. We designed, in collaboration with the staff of Tempo Reale, the interaction mechanisms for one of the main actors of the work, "Nino", played by David Moss. Nino is a character with a double personality: one intelligent and, let's say, crazy and aggressive, the other a sort of repository of knowledge, wisdom, and calmness. These two personalities were modeled by two different live-electronics contexts in his own voice, that he controlled by his movement. David Moss, according to his movement from one part of the stage to the other, had the possibility to change smoothly in real time his voice from one character to the other. David had small sensors on his costume (FSRs), connected to our Wireless-Sensor-to-Midi system (see http://musart.dist.unige.it). Further, a camera placed over the stage was used to detect his overall movement on the stage, to recognize how he occupied the stage. We, therefore, developed a setup able to partially cover both kinesphere and general space. This information was used to change the degrees of intervention of the live electronics between the two contexts corresponding to the two characters. Therefore, he had the possibility to manage a sort of morphing between the two different characters, creating an interesting dialogue between the two Nino personalities.

The EyesWeb system was also recently used in two concerts for two dancers (Opera House of Genova Carlo Felice, May 1998; Civica Scuola di Musica, Milan, June 1998, music by Riccardo Dapelo and Giuliano Palmieri), in an art installation (an active space with a navigating robot at "Arti Visive 2", Palazzo Ducale, oct-nov 1998, Genova, music by Riccardo Dapelo), in a concert of Tempo Reale (Firenze, December 21, 1998, music by Giovanna Natalini), and in a course at Civica Scuola di Musica of Milan (May 1999, with the composer Giovanni Cospito).

CONCLUSION

Music content processing is of growing interest for a number of emerging musical as well as non-musical application areas. Two ongoing projects at our laboratory — EyesWeb and Music Atelier — have been shown to be flexible and reliable enough in a number of other emerging novel application niches, from edutainment to entertainment, rehabilitation, etc. We presented a view and a perspective on intelligent interactive systems. Our proposal is inspired, among other issues, by Laban's effort theory and by KANSEI information processing. In our view, our approach can be considered as a sort of third way between the research in Japan on "KANSEI information processing" on the one hand and the research on "affective computing" in USA on the other.

ACKNOWLEDGEMENTS

The members of the technical staff of our Lab involved in the EyesWeb and Music Atelier projects include Alessandro Coglio, Paolo Coletta, Matteo Ricchetti, Andrea Ricci, Riccardo Rossi, Riccardo Trocca, and Massimiliano Peri. Our project on the KANSEI analysis of expressive content in movement is a joint project in collaboration with Shuji Hashimoto's Laboratory of Waseda University, Tokyo, Japan. We also thank Kenji Suzuki for his contribution to the project at the "Arti Visive 2" interactive art installation and to the development of models of artificial emotions, the composer Riccardo Dapelo, the choreographer Giovanni Di Cicco (Compagnia Arbalete), and Alvise Vidolin.

REFERENCES

Camurri, A. (1995). Interactive Dance/Music Systems. In *Proc. ICMC-95* (pp. 245–252). Banff.

Camurri, A. (Ed.) 1997. *KANSEI: The Technology of Emotion – AIMI International Workshop, Proceedings*. DIST-University of Genova, Italy.

Camurri, A., and Coglio, A. (1998). An architecture for emotional agents. *IEEE MULTIMEDIA*, 5(4), Oct–Dec 1998, 24–33, IEEE Computer Society Press.

Camurri, A., and Ferrentino, P. (1999). Interactive environments for music and multimedia. *Multimedia Systems*.

Camurri, A., and Leman, M. (1997). Gestalt-based composition and performance in multimodal environments. In Leman M.(Ed.), *Music, Gestalt, and Computing* (pp. 495–508). Springer.

Camurri, A., Morasso, P., Tagliasco, V., and Zaccaria, R. (1986). Dance and movement notation. In Morasso, P. and Tagliasco, V. (Eds.), *Human Movement Understanding* (pp. 85–124). North Holland.

Camurri, A., Canepa, C., Orlich, F., and Zaccaria, R. (1987). Interactions between music and movement: a system for music generation from 3D animations. In *Proc. Fourth International Conference on Event Perception and Action*. Trieste.

Camurri, A., Frixione, M., and Innocenti, C. (1994). A cognitive model and a knowledge representation architecture for music and multimedia. *Interface – Journal of New Music Research*, 23(4), 317–347.

Camurri, A., Catorcini, A., Innocenti, C., and Massari, A. (1995). Music and multimedia knowledge representation and reasoning: the HARP system. *Computer Music Journal*, 18(2), 34–58.

Camurri, A., Hashimoto, S., Ricchetti, M., Trocca, R., Suzuki, K., and Volpe, G. (in press). EyesWeb B Toward gesture and affect recognition in dance/music interactive systems. *Computer Music Journal*.

Chadabe, J. (1997). *Electronic Sound*. Prentice Hall.

Dannenberg, R., and Camurri, A. (1994). Computer generated music and multimedia computing. In *Proc. First IEEE Intl. Conf. on Multimedia Computing and Systems ICMCS-94*. May 1994, Boston.

Dannenberg, R., and De Poli, G. (1998). Special issue on synthesis of performance nuances. *Journal of New Music Research* 27(3) (pp. 191–193).

Eshkol, N., and Wachman, A. (1958). *Movement Notation*. London: Weidenfeld & Nicholson.

Hashimoto, S. (1997). KANSEI as the third target of information processing and related topics in Japan. In *Proc. of Intl. Workshop on Kansei – Technology of emotion* (pp. 101–104).

Laban, R. (1963). *Modern Educational Dance*. London: Macdonald & Evans Ltd.

Laban, R., and Lawrence, F.C. (1947). *Effort*. London: Macdonald & Evans Ltd.

Laurel, B. (1993). *Computers as Theatre*. Addison Wesley.

Machover, T., and Chung, J. (1989). Hyperinstruments: musically intelligent and interactive performance and creativity systems. Proc. ICMC'89, 186–190.

Otheru, S., and Hashimoto, S. (1992). A new approach to music through vision. In *Understanding Music with AI*. AAAI Press.

Riecken, D. (Ed.) (1994). *Communications of the ACM*. Special issue on Intelligent Agents, 37(7).

Rowe, R. (1993). *Interactive Music Systems*. MIT Press.

Sloman, A. (1998). Damasio, Descartes, alarms and metamanagement. In *Proc. IEEE Intl. Conf. on Systems Man and Cybernetics SMC'98*. San Diego, CA: IEEE CS Press.

Suzuki, K., Camurri, A., Ferrentino, P., and Hashimoto, S. (1998). Intelligent agent system for human-robot interaction through artificial emotion. In *Proc. IEEE Intl. Conf. On Systems Man and Cybernetics SMC'98*. San Diego, CA: IEEE CS Press.

Vidolin, A. (1997). Musical interpretation and signal processing. In C. Roads, S.T. Pope, A. Piccialli, and G. De Poli (Eds.), *Musical Signal Processing*. Lisse: Swets & Zeitlinger.

Antonio Camurri
Laboratorio di Informatica Musicale
DIST – Department of Informatics, Systems, and Telematics of the University of Genova
Viale Causa 13
I-16145, Genova, Italy
E-mail: music@dist.unige.it
URL: http://musart.dist.unige.it

Antonio Camurri is an Associate Professor at University of Genova where he teaches the courses of "Software engineering" and "Multimedia Systems" (Computer Enineering curriculum). He is founder and scientific director of the Laboratorio di Informatica Musicale, DIST, member of the Board of Directors of AIMI (Italian Association for Musical Informatics), and Associate Editor of the International Journal of New Music Research (Swets and Zeitlinger). His research interests include musical informatics, modeling of expressive content in music and human movement, kansei information processing and artificial emotions, emotional agent architectures, multisensory interfaces and interactive systems. He is author of more than 50 scientific international publications, and served in the scientific committees of several international conferences.

Copyright of Journal of New Music Research is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.