Movement and Gesture in Intelligent Interactive Music Systems

Antonio Camurri
Laboratorio di Informatica Musicale
DIST, University of Genoa
music@dist.unige.it

Riccardo Trocca
Laboratorio di Informatica Musicale
DIST, University of Genoa
ghost@musart.dist.unige.it

This paper introduces a research area on music and interaction by discussing issues on the integration of movement and music languages. As concerns movement, a special focus is given to high-level and qualitative analysis of human movement, starting from previous studies on movement from this viewpoint, such as Laban's Theory of Effort (Laban and Lawrence 1947), and from KANSEI Information Processing (Camurri 1997; Hashimoto 1997). Typical application scenarios include interactive music, theatre, art and museum installations possibly including robots on stage.

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, music, as well as visual and gesture languages, mobile (robotic) scenographies are the components of an integrated multi-level communication. We present this research area and some recent developments by introducing our *EyesWeb* project and some music projects in which these issues have been faced.

The EyesWeb project goals are twofold: from one hand, it aims at exploring and developing models of interaction by extending music language toward gesture and visual languages, with a particular focus on analysing expressive content in gesture and movement, and generating expressive outputs. For example, in EyesWeb we aim at developing methods able to distinguish the different expressive content from two instances of the same movement pattern, e.g., two performances of the same dance fragment. From the other hand, EyesWeb includes the development of an open, visual software for the real-time analysis of full-body movement and gesture of one or more humans and its use in the scenarios described above.

Our research is also inspired by the well-known languages for choreography Laban Theory of Effort (Laban and Lawrence 1947) and Eschol-Wachman and from research in KANSEI Information processing (Hashimoto 1997).

We are successfully employing our research in EyesWeb in various music, interactive art and museal applications, shortly surveyed in the final part of the paper.

State of the art

A well-known research area concerns "hyper-instruments", i.e. virtual musical instruments (e.g., Machover and Chung 1989). Another point of view concerns the development of interaction metaphors which go beyond the "musical instrument", which involve full body movement and gesture at different levels of abstraction, but, above all, concerns the interaction with an active environment capable of evolution, dialogue and not merely a fixed, passive object like an instrument.

In the electronic music community, composers and researchers already faced the problem of integrating dance and music languages, in the framework of the "instrument" metaphor. Already in 1965, John Cage and Merce Cunningham used theremins to capture the movements of dancers in *Variation V*. Several other researchers explored different sensor technologies in the 70's and 80's, mainly to create invisible musical instruments in "active spaces". Other interesting examples are available in the historical overview in (Chadabe, 1997: Chapter 8).

Our research started in 1984 with experiments on gesture and music involving the movement of dancers as a component of the composition. We developed a composition environment allowing the integration of dance and gesture in the music composition process (Camurri et al 1986, 1987). Our focus in this project is on the analysis of the "style of movement" of the dancer: the metaphor is that one of an observer of the dancer(s), able to recognize dance "gestalts". Our software system MANI (Music and ANimation Interface - "mani" means "hands" in Italian), developed in 1986-87, for composers to simulate on a computer a (wireframe, three-dimensional) dancer, to program a dance pattern defining key postures (with the program automatically interpolating between them), and to link a set of "virtual sensors" to animate a virtual wireframe dancer. Virtual sensors in MANI are able to detect parameters of movement, including qualitative, overall (the whole body or parts of it) or gestalt parameters on dancer's movement: for example, MANI included a library of "virtual sensors" able to detect a sort of overall energy, the rhythm, the equilibrium parameters related to a virtual dancer. Our approach was in some aspects similar to the work of Bruno Spoerri based on Rokeby's Very Nervous System in early 90's or recent work on "interactive music" by Todd Winckler.

The parameters detected by MANI virtual sensors are linked to the p-parameters of cmusic score skeleton. This means that a computer simulation of one or more dancing figures generates a *cmusic* score automatically, filled by all the movement parameters related to the *note* statements.

MANI is a software environment simulating an integrated dance-music composition, or, from another point of view, allowing the composer to think of music processes in terms of the metaphor of three-dimensional movement of virtual actors (kinematics chains, not necessarily human-like) and in general of "animated scores".

The next step we made concerned the use of real dancers to investigate the relations between music and dance. We therefore developed another system based on our HARP compositional environment (Camurri et al 1991, 1994), based on special purpose sensor systems (e.g., Costel, MacReflex) to movement detection originally designed for bio-engineering. These allowed a more precise and robust movement analysis with respect to the existing systems (e.g., STEIM BigEye and the Very Nervous System), at the price of higher cost and far more complicate calibration. Versions at lower cost and complexity including proprietary sensor systems were then developed (Camurri 1995, 1997).

Nevertheless, several other systems and approaches to movement analysis have been proposed in the literature, including the very interesting systems developed at Waseda University (Otheru and Hashimoto 1992), with their approach focusing on "KANSEI" (Hashimoto 1997) and the commercial systems used to create believable computer animation in film industry.

Other systems are based on the work of the choreographer Rudolf Laban (e.g., Denney, Lovell and Mitchell 1996) to approach the understanding of human movement.

These two last examples, introducing KANSEI and coreography theories respectively, are two key concepts on which our work is grounded.

Research issues

Multimodal user interaction, artificial emotions, analysis and communication of expressive content are intriguing research issues for novel approaches to interactive systems. Intelligent multimodal interfaces address different human senses 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 from one hand, and visual and gesture objects from the other, are particulartly interesting research topics addressed in several composition and performance environments.

The Emotional Agent Architecture

In (Camurri and Coglio 1998), we propose a general, flexible, and powerful architecture to build emotional agents, i.e. software agents embedding artificial emotions. An 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. 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. The importance of artificial emotions in music content processing is also evident in the field of analysis and synthesis of expressive intentions in musical performances (e.g., De Poli and Dannenberg, 1998).

Research on "affective computing" (Picard 1997) and on "KANSEI information processing" (Hashimoto 1997; Camurri and Hashimoto) concerns research on agents with personality and emotions, from different cultural perspectives. Our proposal, as it will be explained in the paper, is a sort of third way.

In this paper we consider the problem of the *expression* of emotional content rather then 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, output processing (generate sound, music, visual media, control actuators). The external world can be a performance stage, a museal exhibit – real, virtual, or a mix of both – populated by artificial as well as human agents able to interact, dialogue, cooperate, compete, etc. each contributing to the performance. In (Camurri and Coglio 1998) we describe these five components in detail, as well as the data through which they interact with each other and with the external world. We remark that this architecture only provides conceptual requirements for such components and data, without constraining their concrete realizations, which can indeed widely vary across different agent implementations.

Computational realization of an expressive or emotional state, among the many possible, consists in coordinates of a point in some "emotional space". The coordinates can change step by step according to emotional (possibly music) 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"). So, 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 realizations can be found in (Camurri and Ferrentino 1999) and (Camurri et al 1997).

The main role of the emotional component is to influence the processing of the other components by means of parameters containing information about the emotional state (e.g. the name of the zone where the mass is, or a fuzzy vector of point coordinate membership to zones), updated by the emotional component as its state changes.

It is worth to note that the emotional component is tightly integrated with the other components and in particular the rational component. The model includes processes between components which govern the dynamics of expressive behavior. This is also coherent with Hashimoto's proposal of model of KANSEI (1997), who describes KANSEI as a "process" rather then a "state".

KANSEI Information Processing

KANSEI Information Processing (KIP) is a new branch of Information Processing Technology born in Japan. The concepts behind KIP are strongly tied to Japanese culture and way of thinking. KANSEI is a Japanese word that does not have a direct translation in English, or, however, every 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 resolve problems and process information in a faster and personal way. In every action performed by a human being, it can be noticed traces of his/her KANSEI, of his/her way of thinking, of resolving problems, of his/her personality.

Therefore, KANSEI is not a synonym for emotion, although it can be related to emotion, but refers to the human ability of processing information in ways not just logical (Hashimoto 1997). KANSEI is related both to problem solving tasks and to information analysis and synthesis.

An artist expresses, through his works or performances, 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 that will see the work of the dancer, of the actor, the masterpiece of an artist, will use his KANSEI in order to evaluate it, to extract meaning and information about the perceived KANSEI. Basically it is possible to summarize one of the aspects on KANSEI Information Processing 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 that will perceive the results of those actions will use his KANSEI ability to also extract details about the solutions adopted by the performer obtaining details, hints about the personality and the KANSEI of the performer.

2.3. Human Movement Analysis: Laban Theory of Movement

Rudolf Laban was a European choreographer born in 1879 that devoted his life to the study of human movement in all its aspects, ranging from improvement of the performance of workers to the refinement of the expressive power of the movements of dancers and actors. He developed a method to describe movement that was called "Theory of Effort". Laban asserted that effort is a quality of every movement and that the effort present in a gesture and the variations it endures during a movement performance convey the expressive information carried by the sequence of gestures. The work of Laban was 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. According to Laban, expressive power is not only a feature of the kind of gesture performed (a punch, a caress) but it is mainly related to the way of performing it. The variations of some movements factors concur in giving a gesture its expressive power. Laban studied those factors and he found four of them that are collectively named "Effort". It is also possible to think about the 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 the work of Laban.

Laban's theory of Effort can be applied at three different levels of complexity. In its basic form the parameters that describes the effort pertaining to a gesture are four and they can assume just to values. We can refer to this level as "Basic Effort". The four parameters and the values they can assume are shown in table 1.

Factor	Value 1	Value 2
Time	Sustained	Quick
Space	Flexible	Direct
Weight	Light	Strong
Flow	Fluent	Bound

Table 1. Laban Effort components.

Movements performed with "Sustained Time" are movements whose duration is perceived as long and that last long in time with the same characteristics. Pressing against an heavy object in order to move it is an example of movement with sustained time. "Quick Time" is referred to quick and sudden movements, as the reaction in touching something that is really hot. The space factor defines a movement whose path in space is either direct, follows a straight line ("Direct Space") or 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 the eight basic efforts, that are the eight possible combinations of the first three factors (table 2). During a sequence of movements the effort vector describing it moves in a space called "Effort Space" assuming one of the possible states.

	Time	Space	Weight
Slashing	Quick	Flexible	Strong
Gliding	Sustained	Direct	Light
Pressing	Sustained	Direct	Strong
Flicking	Quick	Flexible	Light
Wringing	Sustained	Flexible	Strong
Dabbing	Quick	Direct	Light
Punching	Quick	Direct	Strong
Floating	Sustained	Flexible	Light

Table 2. Eight Basic Efforts.

The second level of complexity is called "Secondary and Stressed Effort" and takes into account of the fact that when performing certain movements, one component of the effort may become predominant, more stressed in respect to the other, which becomes secondary. The basic model is extended and the values assumed by the effort components are no more binary. 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. As seen, in describing effort space, Laban almost neglected the Flux parameter, because it is difficult to directly evaluate, being more complex than the previous three. Furthermore the Weight parameter too can be difficult to evaluate. Usually the level of strength applied in a movement can be appreciated noticing the modifications that the movements exerts on the surrounding environment, for example measuring how much a weight is lifted or shifted, but, in the case of free body movements the performer must mimic the use of strength through the use of appropriate muscles' counter tensions. Weight is therefore an uneasy parameter to evaluate directly, specifically using an automatic system.

Our model of Effort

We developed a model of Effort Space that is useful for a better understanding of effort theory.

In Figure 1 a graphical representation of our model is depicted. The model takes into account just two effort parameters, Space and Time. However it can be easily 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 in various regions:

(1) regions where the characteristics of effort are not well perceivable,

(2) regions where just one if 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),

(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. It is however useful in order to conduct some basic experiment and to go toward building a method to evaluate human movement that is both quantitative and qualitative.

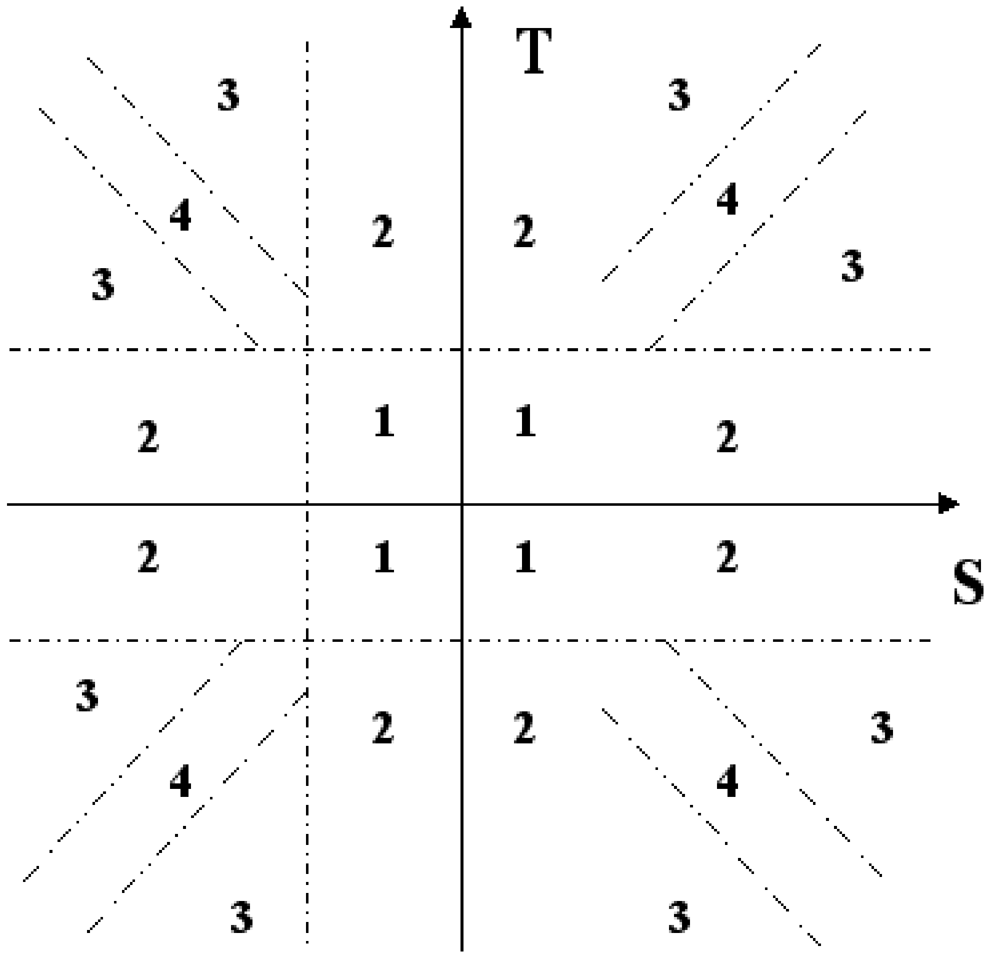

Fig. 1. Graphical representation of the Effort Space model.

Active Spaces and Interaction Metaphors

Gesture intervenes in different aspects and levels in music performance: the most intuitive concerns the simplest causal relationships typical of music (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 (pushing of a switch or an analog sensor) used to produce a note might be not 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 sub-events), the past events in the performance (e.g., to evaluate the influence of preceding gesture on the current event), the tension in the movement toward the next events are of crucial importance in the design of the interaction.

Expectation and anticipation[1]

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

1. I owe Alvise Vidolin interesting discussions and contribute to the clarification of these issues.

Therefore, a deeper analysis of movement is needed to overcome naive uses of interaction mechanisms which often can be noticed in art installations and hyperinstruments. A structured analysis can help in creating anticipation mechanisms, contextual response from a gesture by means of dynamic control. As a very simple example, we recently used floor tiles on-stage for a music production. Usually, digital or analog floor tiles are used, so, when pressed, you can obtain either a 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's FSRs), and a proximity sensor (analogous to Theremin). In this way, the composer receives the information on the velocity, the distance from the tile (by means of analysis algorithms) in order to control the event of "hitting the floor tile" with rich information on its anticipation, including violation of expectations (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 generalised and structured situations in active spaces, also considering the previously described concepts of KANSEI and Effort as 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 to completely re-conceive the VE, the shape and appearance of objects, the way of manipulating and interacting with them in a completely different way, with respect to the 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, i.e. musical instruments.

It is important to mention emerging metaphors which move from hyperinstruments to a sort of dialogue between performers/dancers and the active space designed as their counterpart. For example, if the dancer changes his "style" of movement, say, by gradually reducing the speed of a harsh movement toward smoother gesture, the active space/agent will adapt by continuously changing its behaviour toward a different *context* (*continuously* and in a time interval proportional to the amount of change of the dancer style of movement). Transformations can mean a continuous change both in the sensitivity, interpretation of gesture as well as in the focus of attention, i.e. a change of the set of movements and gestures observed by the system, and 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 of 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* or *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 the sound and music knowledge, the compositional goals, the aspects of integration between music and gesture, including a model of interpretation of gestures and their integration with music parameters, and decides the amount of (possible) degrees of freedom left to the agent as concerns the generative and compositional choices. This is directed toward extensions of music language with action, gesture languages, and possibly visual and animation languages.

Implications in Music Composition and Performance

The discussion in the previous section raises interesting issues about new perspectives on the integration of music and movement languages. For example, the director of the performance in a live electronics (Vidolin 1997) in this scenario has to redefine his role and co-operate with dancers during the performance.

We can easily conceive more sophisticated examples of interaction in which the dialogue between a dancer and the active space can evolve, basing for example on learning, imitation, dialogue: the interaction metaphor can therefore vary from "musical instrument" to "orchestra conduction", to different kinds of "social interaction" under the control of the composer. The role of the composer is now to define the "character" and the "learning capabilities" of the agent the dancer interact with, rather then writing a pre-defined score.

The composer defines in this way the music objects in active terms, deciding which degrees of freedom he leaves to the performer, and for which kind of interaction and evolution of the music object.

It is worth to note that this does not imply necessarily degrees of "randomness" or stochastic processes in the interaction process.

Interesting perspective from the point of view of the director of the performance in live electronics also arise.

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 is already faced in the literature (e.g. Riecken 1994).

In music composition, our hypothesis - and one of the motivations to 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 term 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 in 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) and developed in real time systems (Camurri 1995; Camurri and Leman 1997; Camurri and Ferrentino 1999), and experimented with composers and choreographers in public events.

The EyesWeb project

The general domain scenario include dancers on stage, performers in a live electronics environment, actors in a theatre performance, or the public in a museum or art installation. The objective is to explore extensions of the music language by including dance and, possibly, visual languages.

The Approach

The goal is to support composition processes and interactive performance involving dance and music. The EyesWeb project aims at developing and experimenting approaches and systems able to analyse in real time movement and gesture of one or more humans, with a particular focus to affect and expression in communication.

EyesWeb current hardware architecture is based on multiple b&w, infrared, and color cameras, special electronics, and real-time multi-layered analysis software.

In brief, our approach differs from most existing ones in the following aspects:

- our main focus is not only toward symbol recognition from a posture or gesture vocabulary, but toward the extraction of high-level parameters on expressive intentions in the performance. Existing systems are rather limited from this point of view. Ideally, we want a system able to support the design of an application to distinguish the different expressive intentions from two performances of the same dance fragment. Our focus is therefore on "KANSEI" (Camurri 1997), i.e. on affect and expressive communication;

- the interaction metaphors and the requirements – as discussed in a previous section - and therefore the system architecture, are quite different from existing systems;

- we want to extract both two-dimensional and three-dimensional information about movement;

- to this aim, EyesWeb uses multiple video cameras and other wireless sensor systems, but is not based on the classical stereo vision approach. A typical on-stage configuration with two cameras consists of frontal and lateral views and possibly special localization hardware.

Our system studies human movement from several different points of view and in particular we aim at analysing *effort* in movement (Laban and Lawrence 1947). We work closely with a professional choreographer and dance teacher (Giovanni Di Cicco). In this section we introduce some of the concepts we use in our exploration of human movement.

A *microdance* is 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 we created a set of 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.

Some of the key concepts we are using in our exploration of human motion intention are taken from Rudolf Laban's work. The basic concept of Laban's Theory is Effort, described in a previous section. We model *Effort* as a vector of parameters that identifies the quality of a movement performance. The main focus is on describing the quality of movement. Variations of effort during the movement performance are of particular interest to understand movement. Effort can be applied to the analysis of single parte 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 of 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*.

Laban deems that a movement performance is composed by a sequence of movements executed in 8 main directions generating from the center of the body: forward (f), backward (b), left (l), right (r), left forward (lf), left backward (lb), right forward (rf) and right backward (rb). Every movement sequence is characterized by two factors: the shape created by stretches of spaces and the rhythms created by stretches of time.

The EyesWeb experimental platform

The EyesWeb experimental platform for the study and development of our approach consists of a layered hardware and software architecture. The software 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 to analog synthesizers. Basic libraries include:

- Movement capture modules: input modules for different sensor systems: environmental sensors (e.g. video cameras) and wireless on-body sensors (e.g. accelerometers);
- Filters: low-level filters (e.g. preprocessing, denoise, etc.), high-level filters (e.g. the module to extract the barycenter coordinates of a human figure; the module for evaluating equilibrium);
- Active modules: they have an internal dynamics, they receive inputs as any other kind of modules 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): modules or patches able to extract high-level information, e.g. expressive content, intentions;
- Output modules: low-level (e.g. MIDI, network, DMX, etc) and high-level output modules (e.g. synthesis of expressive intentions able to modulate a score performed in real-time via MAX).

The choice of the significant and reliable module outputs in a given movement and music context – that is, how to dynamically adapt a patch - is a crucial issue in interaction (Camurri and Ferrentino 1999; Camurri and Leman 1997). It usually depends on evaluation of observers outputs, which causes proper feedbacks in patches to activate/deactivate modules and tune parameters.

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

New modules can be created in a simple way by means of a "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 the input and output data types of the new modules, makes available to the user a code area to put user's source code and automatically generates the COM and surrounding code necessary to transform it in an EyesWeb module. EyesWeb is based on the MS COM/DCOM standard. Details on such standard are hidden by the Wizard to the programmer of a module, by generating automatically the necessary COM code.

Filters

The low-level parameters and the processed movement data buffers are directly available to other modules. In this section we describe some details of the filter modules based on the videocamera (IR and color) sensor systems. They are layered on the MIL (Matrox Imaging Library) software for real-time image acquisition and processing.

The low-level modules store frame image buffers of color pixels for each camera. For each frame, they filter noise, extracts the occupation rectangles for each human, extract for each figure a set of low-level parameters (see below).

The tasks of the low-level processing library roughly consists of the following modules, which can be applied to each input channel/videocamera:

- background acquisition, filtering (denoise), and storing as an output for other modules. This task can be invoked initially and at any time as needed by the external in the patch.
- system calibration on the specific dancers low-level, physical parameters (e.g., maximum extension of the open arms of a specific dancer);
- two-level quantization resulting from the difference between the current image and the background;
- denoise of the resulting two-level image with a threshold based algorithm;
- trace the sub-image(s) containing the whole human body figure(s) and compute parameters;
- compute parameters for each rectangle: percentage of black pixels, processing time, equilibrium, main and relative barycenter vectors, etc. (see text below).

This level also includes the processing of possible markers on the body. Both color and digital image buffers and parameters for each dancer are available as outputs from these modules. Examples of available parameters are the barycentres of the 2D silhouettes of the dancers for each camera and peripheral barycentres (ideally, shoulder, elbow, hand, knee, ankle). Taking these information from a frontal and lateral camera, a three dimensional approximation can be computed. For each camera, we have 10 vectors with origin in the main barycentre, whose modules and phases are available.

Modules computing 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) are also available.

As an example of high-level filter, the "equilibrium detector" is defined in terms of the displacement of the lower part of the body near the floor (not necessarity the legs and feet), and the distance of the x-coordinate of the barycenter of the body with respect to 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 spread are the feet (or the parts of the body on the floor) and the more strongly placed on the floor, the more difficult it will be to have sudden and fast full-body movements in the immediate future.

Observers

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. To this aim, the following two issues are particularly relevant:

- the knowledge of the physical laws of human movement (e.g., kinematics 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;
- computational models of artificial emotions can be used in the evaluation of the expressive content of movement, to build interpretations from the data analysed 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 the past stimuli, future expectations, and the agent internal state. We developed algorithms based on artificial emotion models (Camurri and Ferrentino 1999) and 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.

Patch libraries

We developed several patches available with the system. They can be classified in 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 backpropagation 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 to a different camera and, for each frame, to 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. From the point of view of this O, the "tendency to movement" is a function of the distance between feet and of the symmetry of the body barycentre.

The *Microdances observer* uses color videocameras. The 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 match them with a model of typical movement velocity curve extracting the parameters necessary to fit the prototypical curve to the observed data. 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. The 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 it has been shown in section 2 (see the space-time map in figure 1). The patches deal with both movement of a single joint in the kinesphere and the kinesphere in the general space.

Other patch aims at evaluating mobility, stability, lability of movement. An observer module allows to map in real-time a graphical representation of the instantaneous state (or "style") of the movement of the dancer through an animated sequence of points moving on a 2d space.

Fig. 2 and 3. public performance of an EyesWeb patch.

The pictures in figures 2 and 3 shows the use of an EyesWeb patch in a public performance with a robot in an art installation with the public (Arti Visive 2 Museal Exhibition, Palazzo Ducale, Genoa). The robot is itself a visitor of the exhibition and embeds a model of artificial emotions, which is influenced by the stimuli it receives from the external world. It reacts by generating in real time music, video (a "digital mirror" software able to moulde the video signal acquired by the robot) presenting what the robot sees from its on-board camera mirrored on a large screen deformed by its emotional state.

In figure 4 a view of the test window of a single camera for the Input Module is shown: some of the low-level parameters can be seen graphically.

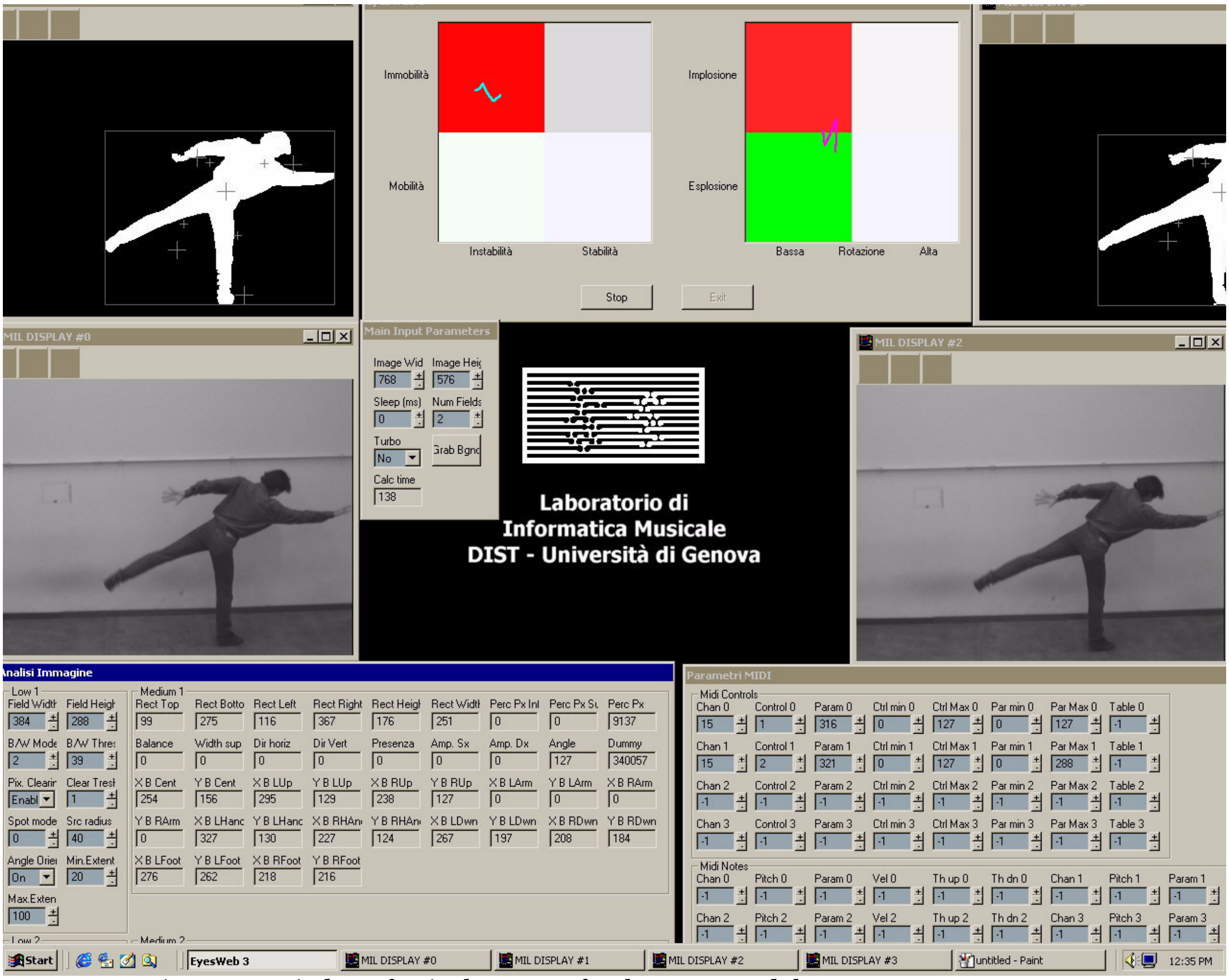

Fig. 4. Test window of a single camera for the Input Module.

4. The Music Atelier museal project at "Città dei Bambini"

We used our systems in several projects. In this section we review the Music Atelier project, dealing with the design of interactive games for science centers. A recent realization of our project is available in *Città dei bambini* (literally "children's city") a permanent science center in Genoa (Italy), where children and kids, by playing interactively with various devices and games, learn notions in the fields of physics, music, biology, and many 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 bio-feedback loop helps visitors to learn some basic music concepts (e.g. rhythm, orchestration). Another game, *Cicerone Robot*, is a navigating and speaking (as well as "music performing") robot which plays the role of a special visitor, explaining and interacting with children (eg, how to play with other games), at the same time playing itself with visitors. The robot can behave changing mood and character in response to whatever is happening in the Atelier. For instance, it gets angry if visitors do not play the other games in the way it explained to 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), environmental lights in the Atelier (e.g. when sad the light becomes blue and dim; the more it gets angry the more red the light becomes). The main educational goals of this game are gently presenting Artificial Intelligence concepts, and ridding the idea in the children (and adults) of science fiction robots 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, while in the other two the sound source - vibration of the membrane - is computer-simulated and controlled by sensors), the *Musical String* (explores the nature of musical timbre), 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 colored little fishes on a computer screen instead of notes on a stave).

As for the hardware and software architecture, 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 a game can depend on how previous games were played, in order to achieve a coherence and continuity in visitors' tour in the Atelier.

5. The use of EyesWeb in Luciano Berio's *Cronaca del Luogo*

We recently used EyesWeb in the music theatre work *Cronaca del Luogo* of Luciano Berio for the opening of 1999 Salzburg Festival (July-August 1999). Here we only shortly 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 and wisdom and calmness. These two personalities are modeled by two different live-electronics contexts on his own voice, that he controls by his movement. David Moss, according to his movement from one part to the other of the stage, 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 detects his overall movement on the stage, to recognise how he occupies the stage. We therefore developed a setup able to partially cover both Kinesphere and General space. This information is 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 different characters, creating a interesting dialogue between the two Nino personalities.

The EyesWeb system has been recently experimented also 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 of Giovanna Natalini), and in a course at Civica Scuola di Musica of Milan (May 1999, with the composer Giovanni Cospito).

Conclusion

We presented a personal view on the concepts and mechanisms of interaction and interactive systems. Our approach developed in EyesWeb has been presented and discussed. Our proposal is inspired, among other issues, to Laban's Effort Theory and KANSEI. In this view, our approach can be considered as a sort of third way between research in Japan on "KANSEI information processing" from one hand, and research on "Affective computing" in USA from the other hand.

Acknowledgements

The members of the staff of our Lab involved in the EyesWeb and music and museal projects include Paolo Coletta, Matteo Ricchetti, Andrea Ricci, Riccardo Rossi, and Massimiliano Peri as well as the composer Riccardo Dapelo and the choreographer Giovanni Di Cicco.

The EyesWeb project become since the last year part of a project on KANSEI Information Processing with Shuji Hashimoto and his Laboratory at Waseda University, Tokyo. His student Kenji Suzuki contributed to the project at the "Arti Visive 2" interactive art installation, concerning the development of the model of expressive communication for the robot. We thank Alvise Vidolin for interesting discussions.

References

Camurri, A. 1995. "Interactive Dance/Music Systems." In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 245-252.

———, ed. 1997. *KANSEI: The Technology of Emotion. Proceedings of the AIMI International Workshop*. Genoa: Associazione di Informatica Musicale Italiana, October 3-4.

———, P. Morasso, V. Tagliasco and R. Zaccaria. 1986. Dance and Movement Notation. In *Human Movement Understanding*, Morasso & Tagliasco, eds, North Holland, pp. 85-124.

———, C. Canepa, F. Orlich and R. Zaccaria. 1987. Interactions between Music and Movement: a System for Music Generation from 3D animations. In *Proceedings of the Fourth International Conference on Event Perception and Action*, Trieste.

———, M. Frixione and C. Innocenti. 1994. "A Cognitive Model and a Knowledge Representation Architecture for Music and Multimedia." *Interface - Journal of New Music Research*, 23(4): 317-347.

———, A. Catorcini, C. Innocenti and A. Massari. 1995. "Music and Multimedia Knowledge Representation and Reasoning: the HARP System." *Computer Music Journal*, 18(2): 34-58.

———, and M. Leman. 1997. "Gestalt-Based Composition and Performance in Multimodal Environments." In *Music, Gestalt. and Computing*, M. Leman, ed., Springer, pp. 495-508.

———, and A. Coglio. 1998. "An Architecture for Emotional Agents." In *IEEE MULTIMEDIA*, 5(4), IEEE Computer Society Press, Oct-Dec, 24-33.

———, and P. Ferrentino. 1999. "Interactive Environments for Music and Multimedia." *Multimedia Systems*, Special issue on Audio and Multimedia, ACM & Springer.

———, S. Hashimoto, M. Ricchetti, R. Trocca, K. Suzuki, G. Volpe (In Press). "EyesWeb – Toward Gesture and Affect Recognition in Dance/Music Interactive Systems." *Computer Music Journal*.

Chadabe Joel. 1997. *Electric Sound, the Past and Promise of Electronic Music*. Upper Saddle River, NJ:Prentice Hall.

Dannenberg, R, and A. Camurri. 1994. Computer Generated Music and Multimedia computing, Proc. *First IEEE Intl. Conf. on Multimedia Computing and Systems ICMCS-94*, May 1994, Boston.

———, and G. De Poli. 1998. Special Issue on Synthesis of Performance Nuances. *Interface - Journal of New Music Research*, 27(3).

Denney, J, R.E. Lovell and J. D. Mitchell. 1996. "Virtual Space Harmony: The Value of Rudolf Laban's Theory of Space Harmony as a means to Navigate a Virtual Stage Environment", Institute for Studies in the Arts, Arizona State University.

Eshkol, N., and A. Wachman. 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 KANSEI: The Technology of Emotion - AIMI International Workshop*, pp.101-104.

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

———, and F.C. Lawrence. 1947. *Effort*. London: Macdonald & Evans Ltd.

Machover, T., and J. Chung, 1989. Hyperinstruments: Musically intelligent and interactive performance and creativity systems. In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 186-190.

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

Picard, R. 1997. *Affective Computing*. Cambridge, Mass.: The MIT Press

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

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

Suzuki, K., A. Camurri, P. Ferrentino and S. Hashimoto. 1998. "Intelligent Agent System for Human-Robot Interaction through Artificial Emotion." In *Proceedings of the IEEE International Conference On Systems Man and Cybernetics - SMC'98*, New York: IEEE Computer Society Press, pp.1055-1060.

Vidolin, A. 1997. "Musical interpretation and signal processing." In *Musical Signal Processing*, C. Roads, S.T. Pope, A. Piccialli, G. De Poli, eds. Swets.

Winkler, T. 1998. *Composing Interactive Music*. Cambridge, Mass.: The MIT Press.

