

INTERACTIVE DANCE/MUSIC SYSTEMS

Antonio Camurri

University of Genova, DIST - Dept. of Communication, Computer and System Sciences,
Computer Music Lab, Via Opera Pia 13, I-16145, Genova, Italy

e-mail: music@dist.unige.it

(voice) +39-10-3532988

(fax) +39-10-3532948 / 2154

ASBTRACT: This paper focuses on interactive systems for the real time acquisition of human movement (e.g., dance), sound and music input, for the “intelligent”, high-level control of sound, music, and multimedia. The introductory part of the paper discusses the need for systems able to integrate sound/music and movement/dance at a higher-level than mere synchronization: to this aim, some main research issues and requirements are presented. In the second part, we set out to address the problems and the requirements of such systems, and as a partial response in this respect, examples of interactive dance/music systems, based on HARP hybrid agent architecture, are described.

1. Introduction

As a part of a general investigation on multimodal interaction in multimedia systems in the framework of the Esprit 8579 MIAMI Project, our research is mainly concerned in the study and development of autonomous, multimodal interactive systems (MISs), i.e. systems characterized by multimodal, real-time interaction with user(s), possibly in unstructured and evolving environments. The scenario considered in this paper consists of users immersed in a multimodal experience but not in a conventional virtual environment that he alone can perceive. Rather, we envisage an audio-visual environment which can be communicated to other humans, either other actors participating in the same event or external spectators of the action. We can think of a sound/music/light/image/speech synthesis system which is driven/tuned by the movements of the actor(s) using specific metaphors for reaching, grasping, turning, pushing, navigating, playing, communicating states or emotions, etc.

Axel Mulder (1994) gives an interesting survey and a classification of a subset of such systems, that he calls *virtual musical instruments*: well-known examples are *hyper-instruments* (Machover and Chung 1989), which, similarly to physical musical instruments, allow performers to map gestures and continuous movement into sounds, and *multimodal orchestras*, which can be applied to choreography, dance, music.

Figure 1 shows a general architecture of MISs. Different families of sensors can be adopted, according to the type and the “grain” of the information we need about human movement. On-body sensors are typically useful for relative motion analysis (e.g., distance between hands), trajectories tracking, fine motion detection; external sensors can integrate these information with absolute positions tracking, global movement patterns and “style” extraction (for example, measures of the “coordination”, the amount of “energy”, the volume occupation in the time unit, in general qualitative information integrated over time windows, etc.). In our experience, it is more effective to integrate different simple and robust sensor technologies, rather than to use of a single, even sophisticated, sensor technology. The systems described in this paper are based on different sensor technologies - *V-scope* and *SoundCage* - which can be integrated since the common supervision system kernel. They will be jointly experimented in a concert on summer 1995.

From the point of view of the *supervision software system*, the complexity of the domain - which goes from sound and music knowledge processing, to advanced robotics, intelligent user interfaces, sensor systems, computer and multimedia technology in general - requires “hybrid” models integrating such different representations. To this aim, *integrated agents architectures* demonstrated to be an effective platform for such categories of systems and have been already experimented (Rowe 1993). We speak of “autonomous” multimedia systems, that is, autonomous *agents* characterized by multi-modal interaction with user(s) (as well as between agents themselves) in an unstructured and evolving environment (Ferguson 1992; Steels 1994; Riecken 1994).

In this paper we explore such MISs architectures and we give an overview of a family of systems we are developing. The research described in this paper starts from previous work on the design of systems for the study and integration of dance and music (Camurri et al 1986; 1993; Ungvary et al 1992), and on high-level interactive systems like Cypher (Rowe 1993) and HARP (Camurri et al 1991, 1995). In particular, HARP (*Hybrid Action Representation and Planning*) concerns the design of hybrid agent architectures for the representation and real-time processing of multimedia knowledge, able to support users in the design and multi-modal control and production of music and multimedia material, managed at different levels of abstraction. The current version of the HARP system runs under Windows 95 and as been implemented in MS Visual C++ and Quintus Prolog.

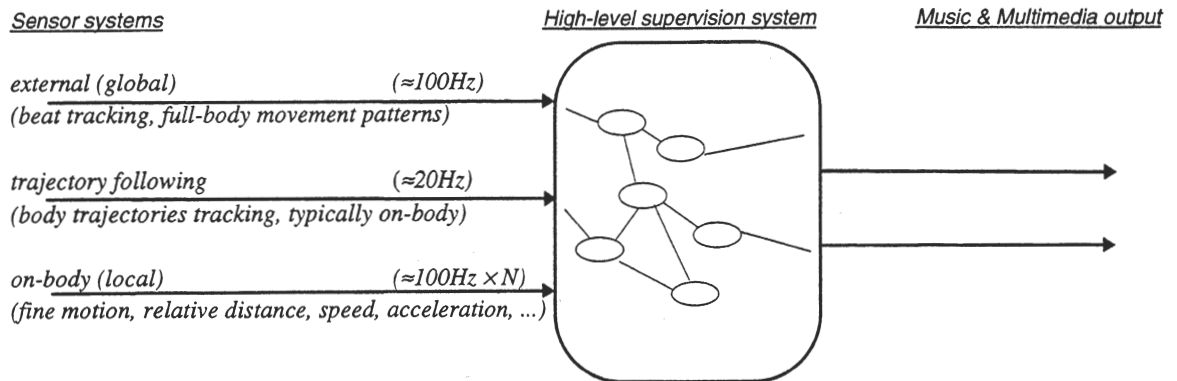


Figure 1. MIS Overall Architecture

1.1 Virtual Reality Systems and Multimodal Interactive Systems

Several applications of VR have been developed for theatre and music (see for example Laurel et al 1994). It is important, at this point, to distinguish between virtual reality (VR) systems (or virtual environments) and MISs. The main difference lies in the intention behind their respective research directions: VR seems to imitate reality to establishing absorbing audio-visual illusions, whereas multimodality attempts to enhance the throughput and the naturalism of man-machine communication. The audio-visual illusion of VEs is merely a trick for triggering the natural synergy between sensory-motor channels that is present in the brain. This is not the only possibility for an effective exploitation of the parallelism and associative nature of human perceptual-motor process: VR research can be considered a subset of research in multimodality. For example, MISs can be viewed as computer-as-tool and computer-as-dialogue-partner, which are further alternatives to the computer-as-audiovisual-illusion typical of VR systems.

2. Basic Requirements of High-Level Multimodal Interactive Systems

In this section we analyse the basic requirements and issues which high-level supervision systems should challenge with. We need formalisms that can manage the multifarious structure and levels of abstraction of music and multimedia objects, from *symbolic* abstract representations (e.g., the symbol evoked by a body gesture) to *subsymbolic* representations (e.g. a 3D trajectory of a part of the body; a signal perceived by the human ear). Different views of the same object are often necessary: according to the reasoning perspective and the goal to be reached, multimedia object representations can vary from an atom in a symbolic high-level representation to a stream of low-level signals in the broadest view of the same material. More specifically in music, Leman (1994) distinguishes three main types of level: (i) the acoustic (signal) level, (ii) the auditory (subsymbolic) level, and (iii) the conceptual (symbolic) level. This can easily be extended to other media, and MISs should be able to represent and manipulate objects at all these levels.

Analogies and metaphors are crucial issues in the representation and reasoning capabilities of MISs. They are widely used in music languages and are at the basis of languages for the integration of different multimedia knowledge. Let us consider the case of music languages: their metaphorical richness derived from real world dynamics is significant - see for example (Camurri et al. 1986). In general, the terms and descriptions in one modality can be used to express intuitively "similar" concepts in others. We consider analogies and metaphors to be the basic "glue" for integrating different modalities, with particular regard to sound/music and movement/dance representations. The issue of reasoning based on metaphors has been widely studied from various AI, psychology and philosophy standpoints. Steps toward an approach to modelling *metaphors* can be found for example in (Gärdenfors 1988): his theory analyzes metaphors in terms of topological structure similarities between dimensions in a *conceptual space*. Diagrammatic or pictorial descriptions of situations (Chandrasekaran et al. 1993) are another interesting field of research in this respect.

Furthermore, formalisms that can support users should provide mechanisms for *reasoning on actions and plans*, and for analyzing alternatives and strategies, starting from user requirements and goals. They should provide both formal and informal analysis capabilities for inspecting the objects represented.

Other points are *learning* and *adaptation*, i.e. how to automatically update or adapt system knowledge to new information and to the nonlinear human behaviour. Several solutions have been proposed in the AI literature, such as purely symbolic approaches and learning systems based on neural networks (see for example Lee and Wessel 1992). As for human movement recognition, some learning problems can be

translated into a dimensionality-reduction problem of extracting "principal components" from a high-dimensional space of redundant degrees of freedom. In any case, we need learning because in multimodal systems there is not, in general, a simple one-to-one translation of signals and events as in VR systems.

3. The HARP Hybrid Architecture

HARP can be considered as a step toward the generation of MISs following the requirements previously discussed. It has been developed for the representation and real-time processing of multimedia knowledge, as a sort of supervisor that supports users in the design and the multimodal control of multimedia material, at different levels of abstraction. The system is able to represent and carry out plans in real time for manipulating knowledge according to the user's goals. HARP is based on a hybrid integrated agent architecture. Here, the term "hybrid" means "combining different formalisms". The overall system architecture is a distributed network of autonomous agents (Steels 1994): from this viewpoint, the system is similar to *Cypher* (Rowe 1993), *TouringMachines* (Ferguson 1992), and *M* (Riecken 1994). The current HARP model and the system are described in more detail in (Camurri and Innocenti 1995). The audio compact disk from IEEE CS that accompanied the July 1991 issue of *Computer* magazine includes several music examples produced with an earlier version of the system (Camurri et al 1991).

3.1 The Symbolic Components

From a cognitive viewpoint, the system is structured in a long-term memory (LTM), the permanent, "encyclopedic" storage of general knowledge, and in a short-term memory (STM), the actual "context" regarding the state of the affairs of the world and the problems currently faced. Both LTM and STM are composed by symbolic and subsymbolic components, to face the different structure and nature of the domain knowledge. The symbolic LTM consists of two components:

1. a *terminological component* appropriate for defining terms and for describing concepts and the taxonomic relationships between them. An inheritance semantic network formalism (Woods and Schmolze 1992) has been extended to represent and reason on time, actions and plans, and encompasses the possibility of using first-order axioms to extend its expressiveness.
2. an *assertional component* to represent factual long-term knowledge on the domain, based on first order logic. For example, the incipit of a well-known piece, say, Beethoven's fifth, is an assertional constant which can be considered part of the LTM. The assertional component can include factual generalizations expressed as first-order axioms (e.g. by means of quantified implications): for example, in the museal or theatrical scenario "all the moving objects on stage are humans".

The symbolic STM contains information concerning the specific events represented in its subsymbolic counterpart (e.g. sensory input). In other words, the STM symbolic component represents a single context of the actions represented in the subsymbolic component: it is a one-to-one symbolic representation of the entities and of the events in the subsymbolic component.

3.2 The Subsymbolic Components

The STM symbolic knowledge base (KB) is linked one-to-one to the STM subsymbolic representation, which can either be connected to the world itself by sensors/actuators or to a mental model of the world.

The two subsymbolic components are the direct counterparts of the symbolic components. They consist of a network of cooperative *agents* in the sense of (Steels 1994): each agent is a class in a concurrent object-oriented system; agents are hooked to terms in the symbolic components.

The generation and activation of a network of agents produces a simulation, an execution, or measurements in the real environment. Low-level sound representations, data on real-time performance, as well as human movement recognition and synthesis algorithms are all examples of classes of agents in such a concurrent environment. Each is "hooked" to proper terms in the symbolic KB.

The definitions of agents are resident in the subsymbolic LTM, in a taxonomy isomorphic to the symbolic LTM. To activate agents, they must firstly be instanced in the subsymbolic STM. Therefore, subsymbolic activity can only be carried out in the STM, i.e. in a completed context. The context can include the sonological level of music objects, as acquired by a model of the auditory system, instances of the transformation processes on such objects possibly depending on human movement/dance patterns. In a theatrical, "generalized choreography" model, the context can also include the agent internal representation and the three-dimensional world, possibly acquired by consulting its sensory input.

3.3 Signal Handling by Agents and Icons: Subsymbolic Reasoning

Two kinds of subsymbolic entities are devoted to the management of symbols and can be instanced to form a context of cooperative entities: *agents* and *icons*. An *agent* is a class which is "expert" in a sub-set of the

domain, and which performs actions upon it. *Icons* encompass two aspects; firstly, diagrammatic or pictorial descriptions of situations (Chandrasekaran et al. 1993). These can be geometrical metaphors of a different domain: for example, several contemporary music notations used by composers (e.g. Kagel, Bussotti, Berio, Ligeti) are based on such kind of metaphors. The second aspect covered by icons is that they act as dynamic systems, as metaphors for reasoning on actions and plans. Landscapes of energy and models based on force fields are simple cases considered in this paper. We can see this kind of representation as an enrichment of the previous one, since it includes dynamics, force, and time coordinates within diagrammatic representations. Different *agents* that can perform navigation algorithms on N-dimensional maps have been defined in the subsymbolic component of the LTM, a useful element in different domains. Force fields can be built by learning processes, as in the case of Leman's TCAD attractor dynamics system (Leman 1994) based on artificial neural networks.

The activity of agents and icons can be interpreted as a reasoning mechanism complementary to typical symbolic deductive systems: for example, navigation in a force field can substitute decision processes which would otherwise be difficult to model symbolically (e.g. with axioms or rules).

The symbolic reasoning differs from subsymbolic reasoning in *time granularity*: subsymbolic reasoning is expected to react in real-time, since it has to follow and manipulate the flow of signals which usually require strict time constraints. Symbolic reasoning is usually expected to intervene at a much higher time granularity (a few seconds vs. milliseconds). For example, during the tracking of human movement, we have a set of agents processing the input signal (including self-organizing networks for input classification); at "significant" instants certain gesture can be recognized and then instanced and therefore manipulated by the symbolic component.

The *context manager* supervises communication between agents and icons. These do not know whom they are communicating with, since they must be general, reusable objects: agents and icons are instanced and connected to each other by the context manager according to the specific context and the particular action/plan to be performed. The context manager defines and updates a set of input and output communication links for each agent and icon in the current context, complying with constraints in the symbolic memory (e.g. the topology of the semantic network and the current STM contents). The subsymbolic memory is therefore built and kept aligned with the symbolic memory during execution.

Significantly, the context manager is able to manage and control the execution of hierarchical actions, i.e., actions recursively expanded into subactions by means of *part of* (or *aggregate*) relations, connected to input/output situations in hierarchies (in terms of both *IS-A* and *part-of* links). The algorithm of the context manager is described in full detail in (Camurri and Innocenti 1995).

4. The HARP/V-Scope System

This section describes a HARP application for the tracking of human movement by means of the *V-Scope* sensor system, and its integration with animated human models and sound and music.

This HARP application is composed of four main subsystems, each corresponding to a group of agents:

- a) a human movement acquisition sub-system based on *V-scope*;
- b) human movement data pre-processing agents;
- c) human movement and gesture recognition agents, including agents based on the force field metaphor;
- d) system output: sound and music, computer animation.

An excerpt of the symbolic KB is shown in figure 2, and a corresponding agent software architecture is shown in figure 3. In figure 2, double arrows are inheritance (*IS-A*) links, single arrows are relations including *aggregation* (*part-of*); relations are defined between concepts (classes) represented by ellipses. This language is derived by KL-ONE (Woods and Schmolze 1992). For example, *vscope_action* *IS-A* *compound_action*, formed by (*part_of* relations) *voice_area*, *bell_area*, and *orchestral*. On the left part of the figure the situations and the significant parts of the body tracked in the experiment (hands and torso) are represented. The symbolic KB shown in figure 2 is completed by the agents and icons linked to situations, actions, and their derived subclasses. Let us analyse in more detail the network of agent instances built by HARP in this experimental setup.

4.1 Subsystem (a) - *V-scope* interface

The *VScope* agent is designed to acquire information on the position of a number of *V-scope* markers typically placed on the body of a user. It manages both low-level serial communication and the link with client modules. *V-scope* is an infrared/ultrasound sensing device developed by Lipman Ltd. for the real-time acquisition of the position of up to eight markers placed on the human body (e.g. on the articulatory joints) or in general on moving objects (e.g. a video camera). The hardware is composed of the markers,

three tx/rx towers for real-time detection of marker position, and a main processing unit connected via a serial link to a computer. The sampling rate can vary from 5 to several hundreds of ms per marker. As for the limits imposed by the V-scope hardware, the stage can vary from 2 to 5 meters in depth: faster sampling corresponds to a smaller area, due to the limitations of the ultrasound sensors. Our experimental results show that 12-15ms per marker is a good tradeoff between speed (a satisfactory value for human movement acquisition without losing too much information) and stage size. The precision of V-scope is in the range of $\pm 0.5\text{cm}$, an acceptable magnitude for our application.

The symbolic knowledge related to the management of the V-scope data is shown in the left part of the KB in figure 2. The icon attached to *vscope_situation* (the *Shared Dynamic Link Library - DLL* in figure 3) makes the sensor system data available to agents. Both low-level methods for the V-scope hardware management and high-level communication methods are encapsulated in the *Shared DLL* icon.

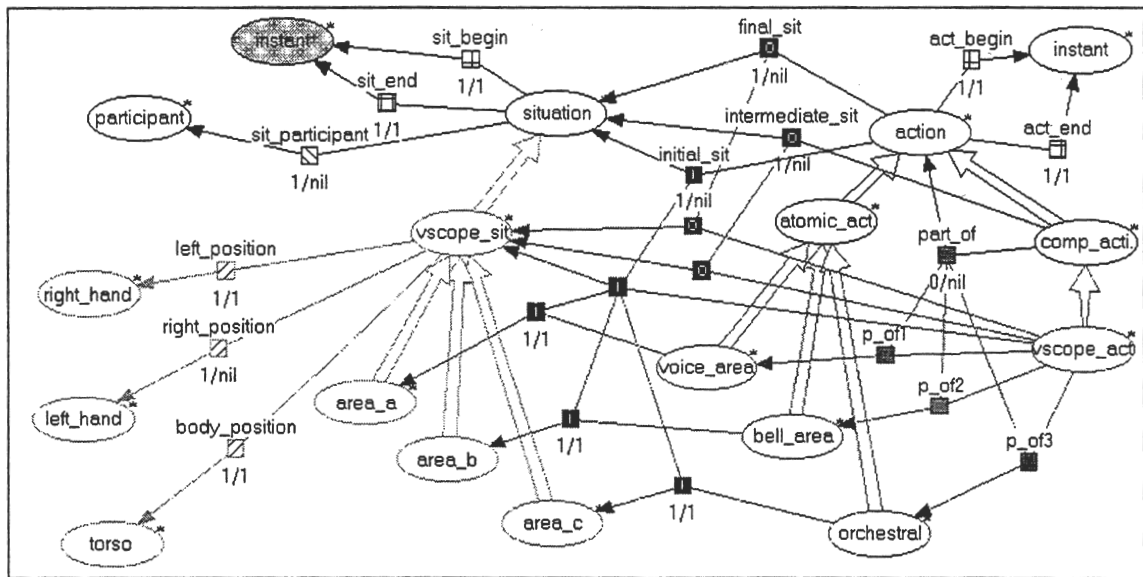


Figure 2 - Excerpt of the symbolic KB of the HARP/V-scope system

4.2 Subsystem (b) - Human movement data pre-processing agents

The motion data preprocessing agent receives as input the raw data stream from the V-scope icon and filters it to ensure no spurious information is present and that values are within a meaningful range.

4.3 Subsystem (c) - Movement and Feature Extraction. The force field metaphor

The communication agent reads the filtered sensors' data stream and forwards it to the gesture recognition and movement analysis agents, whose output can trigger or influence the activities of sound/music and animation agents. A simple example is when a feature agent recognizes that a dancer has raised his/her left arm over a certain threshold, and its output can be used to activate a certain sound processing agent. The gesture recognition task is further subdivided into several concurrent agents, each dedicated to a different task. The Gesture/Movement agents implemented in the experiment described here are able to recognize several different features and gestures: raising and lowering one or both hands, raising and lowering the body, opening and closing the palm of both hands, distance between hands and gesture speed. Another sub-category of Gesture/Movement agents is based on the force field metaphor: for example, we investigated the mapping of (x,y) coordinates of the dancer into force fields. Figure 4 shows the sensorized area and, in the video screen, it is possible to notice an agent window containing a visual description of a force field with three peaks (white blobs) corresponding to significant places in the sensorized area. Each place is characterized by a different behavior, i.e. different mappings of movement/sound.

A further category of agents extract higher-level features from the movement, to contribute to model more complex music/movement correlations. Simple examples of high-level features are "how fast the movement is", "how in tempo the dancer is moving". This kind of information is obtained from integration over time slices. In most cases two different time slice ranges, approximately 0,5-1s and 3-5s, showed the best results. Time slices on which agents operate can vary dynamically, e.g. on the basis of "goodness" of the

recognition: for example, a decrease of the quality in the movement recognition (e.g., different agents return conflicting data) can cause the agents to vary their time granularity and/or their time slice on which they operate. Agents based on self-organizing neural networks for the classification of incoming data from sensors, including acoustic signals have been developed. The experiments involve both the V-scope and the SoundCage agents (described in a next section) to cope with different sensor outputs and conflicts.

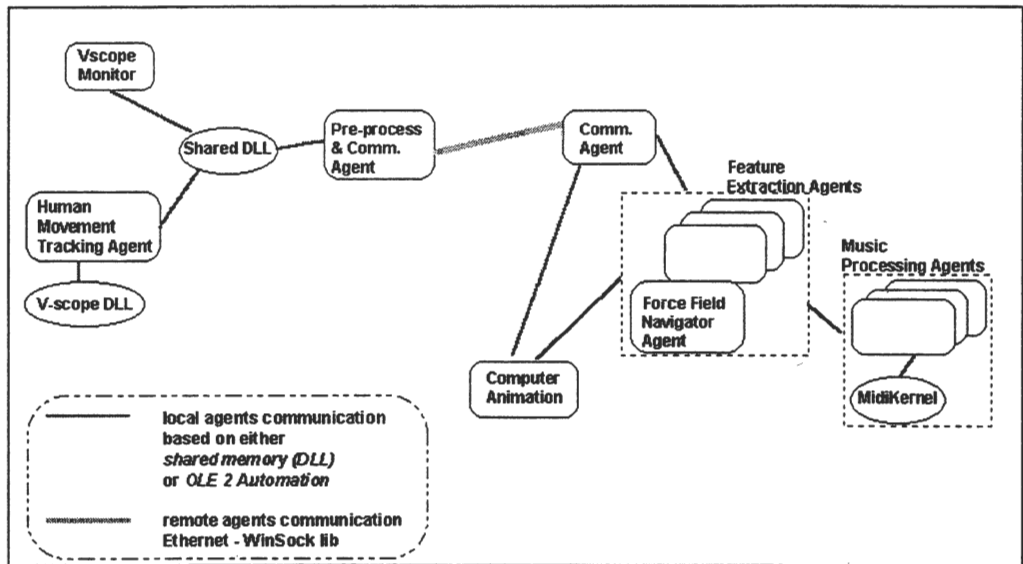


Figure 3 - The HARP/V-scope subsymbolic agents

4.4 Subsystem (d) - Output Generation

In the simplest case, the recognized features can be directly used for the real-time control of music events and computer animation. The mapping of the performer's movement to sound and animation agents can be either pre-defined or dynamically updated according to the information acquired by particular feature extraction agents. MIDI is used for sound event control: sound output agents receive MIDI commands through OLE links from movement recognition or reasoning agents, and queue them on the agent which manages the low level scheduling and synchronization and the sound output to synthesizers.

In the example of figure 4, we use three hyper-instruments corresponding to different areas of the sensorized stage. The three areas/hyper-instruments correspond to peaks in the force field in figure 4: the picture shows HARP/V-scope at work with a user (the computer screen with the force field window indicating the position of the user can be seen). In this experiment we used three markers - one for either hand and a third for generic body location. This last one is useful in capturing (i) body position in the force field map (x and y coords), and (ii) body height position (z coord), e.g. it establishes whether the dancer is standing or crouched. It is possible to recognize different hand gestures for each different hyper-instrument. The markers on the hands can be tracked only if the user keeps the hands open; this is used to control the starting and stopping sound outputs, performed respectively by opening and closing a hand. When the performer is in the center of an area, we obtain maximum presence of that particular instrument, while the other two are absent. As the dancer moves from one peak to another, a cross-fading effect from one instrument to the other is attained: in general, the output of the three instruments is mixed according to the shape of the force field. The sound synthesis techniques controlled in real-time in this demo are formants vowel synthesis and Karplus-Strong, implemented on the IRIS/Bontempi SM1000 hardware.

The HARP/V-scope system has been implemented under Microsoft Windows 95 operating system. A Pentium PC is physically linked to the V-scope hardware via a high-speed serial interface (RS232C with 16550AF UART) and hosts the above described subsystems. A Windows sockets library has been developed to allow the link to possible remote subsystems.

5. SoundCage

A further project regards the use of HARP as a supervisor for *SoundCage*TM, a device for the real-time acquisition, processing of the movements of a dancer, developed by *SoundCage S.r.l.* The SoundCage hardware is composed by a proprietary infrared and pressure sensor systems in a sort of "cage" structure

(the stage) and by special low-cost I/O boards. SoundCage has no on-body sensors. Its output is used to drive MIDI digital synthesizers and other multimedia devices (e.g., lights and special effects, video-wall). The SoundCage *real-time software kernel* is similar to the HARP KB and agents architecture described in figures 2 and 3. Different Movement/Gesture recognition agents are delegated to the acquisition of sensor data, others to internal computation (e.g., for changes in the sound/movement associations). Agents are dynamically connected in a network, including the output agents, which control the MIDI and multimedia output. The SoundCage software includes also a *development environment*, which is used by the composer/coreographer for the off-line phase of creation of music/dance composition/coreographies.



Figure 4 - HARP/V-scope at work. The user position is the black dot in the window on the screen.

6. Theatrical Machines

A different category of MISs is based on the integration of multimedia technology with advanced robotics for the development of systems that can interact with actors on a theatre stage - and possibly with the audience. Theatrical machines are basically dedicated to the supervision and real time control of real autonomous agents on a theatre stage (e.g. a real vehicle on wheels, equipped with on-board sensors, a computer for the low-level processing of sensorial data, etc.). Such systems move, navigate and react to events happening on stage (e.g. actions performed by the actors or the spectators), acquire sounds from the environment, and possibly perform musical tasks. We developed a prototype of such a machine, which includes a workstation that runs HARP - the general supervisor - and a set of software tools (drivers and communication software) for managing the radio link with an autonomous mobile robot (initially a LabMate hardware and electronics, currently a more sophisticated robot). The robot has a PC 486 on board for the low-level processing of sensors and actuators, and for the implementation of low-level reactive navigation algorithms. It is equipped with a set of sound and multimedia devices: the current implementation features a stereo audio amplifier and loudspeakers, and a radio-control for video monitors located in the stage or exhibition area. The architecture also includes a separate radio link for sending audio signals to the robot (voice, sound and music signals produced in real time on the supervision workstation and sent to the robot), and an original positioning system - called Local Positioning System - based on a variation of the techniques of GPS (global positioning system), developed at DIST by Piero Morasso and Renato Zaccaria for use in interiors for the tracking of robot position during navigation.

Thanks to the HARP engine, the system integrates planning capabilities, both symbolic and hybrid. For example, a choreographer can define a particular percouse and the behaviour in areas of interest in the stage, possibly depending on the current context. This information is used by the system to plan a navigation pathway in the stage area, exhibiting different behaviours along its way. Further hybrid planning is performed by the system in several cases: for example, during navigation, if the robot must reach a goal inside an area with a small entrance, the potential field information might not be sufficient to allow it to pass the doorway. So, the symbolic component recognizes this and other cases, and transforms the initial goal into a compound action with an intermediate goal corresponding to the reaching of the doorway (a temporary attractor is put on the doorway), followed by the initial goal. Initial public experimentation of a prototype that included a subset of the above-described architecture was carried out at Palazzo Ducale (Centro dei Dogi, Genova), 19-22 December 1993, in the context of a museum exhibition.

Acknowledgements: *This work has been partially supported by the Esprit Project 8579 -MIAMI (Multimodal Interaction for Advanced Multimedia Interfaces), and by a special project of the Italian National Research Council (CNR) on interactive multimedia systems for art and entertainment. Carlo Innocenti, Marcello Frixione, Marc Leman and Renato Zaccaria made important contributions to the definition of the current HARP cognitive model. Piero Morasso made a vital contribution to the museal project and to human movement representation issues. Alessandro Catorcini and Alberto Massari implemented most of the software modules of the HARP system. Claudio Massucco contributed to the system design and implemented the MIDI library and the real time SoundCage software. Roberto Chiarvetto and Riccardo Rossi implemented most of the HARP IV-Scope application.*

We wish to thank IRIS-Bontempi for the availability of the MARS workstation, and Simon Fraser University for the availability of the LifeForms software license. SoundCage™ has been developed by SoundCage S.r.l.

References

- C.Ames, "AI and Music", *Encyclopedia of Artificial Intelligence*, 2nd ed., New York: John Wiley and Sons, 1992.
- T.W.Calvert, A.Bruderlin, S.Mah, T.Schiphorst, C.Welman, "The evolution of an interface for choreographers", *Proc. INTERCHI 93*.
- A.Camurri, P.Morasso, V.Tagliasco, and R.Zaccaria, "Dance and movement notation", in P.Morasso and V.Tagliasco (Eds.), *Human Movement Understanding*, North Holland, 1986.
- A.Camurri, C.Canepa, M.Frixione, and R.Zaccaria, "HARP: A System for Intelligent Composer's Assistance", *IEEE COMPUTER*, 24(7), July 1991, 64-67. Revised extended version in D.Baggi (Ed.), *Readings in Computer Generated Music*, 95-115, IEEE CS Press, 1992.
- A.Camurri, F.Giuffrida, G.Vercelli, "A system for the real-time control of human models on stage." *Proc. X Colloquium on Musical Informatics*, AIMI and University of Milan, Italy, 1993.
- A. Camurri, C.Innocenti. "A Hybrid System for Music and Multimedia Processing." *DIST Techn. Rep.*, May 1995..
- B.Chandrasekaran, N.Hari Narayanan, Y.Iwasaki, "Reasoning with diagrammatic representations - A Report on the Spring Symposium", *AI Magazine*, Vol.14, No.2, pp.49-56, Summer 1993.
- R.Davis, H.Shrobe, P.Szolovits, "What is a Knowledge Representation?", *AI Magazine*, Vol.14, No.1, Spring 1993.
- I.A.Ferguson, "TouringMachines: Autonomous Agents with Attitudes", *IEEE Computer*, Vol.25, No.5, May 1992.
- P.Gärdenfors, "Semantics, Conceptual Spaces and the Dimensions of Music", *Acta Philosophica Fennica*, Essays on the Philosophy of Music, Rantala, Rowell, and Tarasti (Eds.), Vol.43, pp.9-27, 1988, Helsinki.
- B.Laurel, R.Strickland, R.Tow, "Placeholder: Landscape and Narrative in Virtual Environments," *Computer Graphics*, 28(2), 118-126, May 1994.
- M.Lee, D.Wessel, "Connectionist models for real-time control of synthesis and compositional algorithms". *Proc. ICMC-92*, San Jose, CA, USA, 227-280, 1992.
- M.Leman, "Schema-based tone center recognition of musical signals". *Interface - Journal of New Music Research*, Swets & Zeitlinger, Lisse, The Netherlands, 1994.
- T.Machover, J.Chung, "Hyperinstruments: Musically intelligent and interactive performance and creativity systems." *Proc. Intl. Computer Music Conference*, Columbus, Ohio, USA, ICMA, 1989.
- A.Mulder, "Virtual Musical Instruments: Accessing the Sound Synthesis Universe as a Performer", *Proc. First Brazilian Symp. on Comp. Music, 14th Annual Congr. Brazilian Comp. Soc.*, Caxambu, Minas Geiras, Aug. 1994.
- D.Riecken (Ed.) 1994. Special issue on Intelligent Agents, *Communications of the ACM*, July 1994.
- R.Rowe (1993). *Interactive music systems*. The MIT Press, Cambridge, MA.
- T.Schiphorst, T.Calvert, C.Lee, C.Welman, S.Gaudet, "Tools for Interaction with the Creative Process of Composition", *Proc. CHI '90*, 167-174, 1990.
- L.Steels (1994). The artificial life roots of artificial intelligence, *Artificial Life*, 1(1-2), 75-110, MIT Press.
- T.Ungvary, S.Waters, P.Rajka, "NUNTIUS: A computer system for the interactive composition and analysis of music and dance", *Leonardo*, 25(1), 1992, pp.55-68.
- W.A.Woods, J.G.Schmolze "The KL-ONE family", *Computers Math. Applic*, 23(2-5), 133-177, Pergamon Press, 1992.