EXPRESSIVE GESTURAL CONTROL OF SOUND AND VISUAL OUTPUT IN MULTIMODAL INTERACTIVE SYSTEMS

Antonio Camurri
Antonio.Camurri@unige.it

Barbara Mazzarino
Barbara.Mazzarino@unige.it

Gualtiero Volpe
Gualtiero.Volpe@unige.it

InfoMus Lab – Laboratorio di Informatica Musicale
DIST – University of Genova
Viale Causa 13, I-16145
Genova, Italy
www.infomus.dist.unige.it

ABSTRACT

This paper discusses a multilayered model for strategies for real-time expressive control of sound and visual output. A particular focus is on multimodal interactive systems for performing arts as user-centred systems able to interpret the high-level information conveyed by performers through their expressive gestures and to establish an effective dialog with them taking into account expressive, emotional, affective content. Two kinds of control strategies are discussed enabling the designer of a performance to develop interactive systems able to react to performers' expressive actions, to build a dialog with performers, or to mix purely reactive behaviour with more complex dialogical situations. A further layer of processing allows dynamic adaptation of control strategies to the evolution of the performance. A quick survey of algorithms for expressive gesture processing is also provided. Expressive gestural control strategies are described with reference to current and past research projects at the DIST- InfoMus Lab. Some concrete applications based on the EyesWeb open platform (www.eyesweb.org) are presented as well.

1. INTRODUCTION

Performing arts demonstrated to be a key research and application field for multimodal interactive systems (see for example [23][24]). From the point of view of research, performing arts are an ideal test-bed for works concerning mechanisms for non-verbal communication of affective, emotional, expressive content. For example, in [4] music and dance performances have been employed for studying expressive gestures and their ability to convey emotional states (e.g., the well-know and consolidated basic emotions), and to engage spectators. As for applications, multimodal interactive systems may assume a very relevant role in the design and development of novel systems and tools for performing arts. They allow artists to access to and to use in real-time information about the ongoing performance that it would be impossible to gather without the aid of technology. Applications may range from the development of a novel generation of gesture controlled musical instruments (see for example [1]), toward the definition of completely novel art forms in which technology is not only an infrastructure for an artwork, but rather it is integrated with art at the level of language and it becomes something intrinsic to the artwork.

Concrete examples of such trend can be found in the recent developments of musical informatics in which a growing interest can be noted on research on innovative human-computer interfaces for performing arts. Consider for example the evolution from hyper-instruments [20], i.e., traditional instruments enriched with sensors extending the possibilities for sound control, to virtual instruments, i.e., the space surrounding a performer becomes an instrument sensible to performer's gestures and behaviour (see for example [13]), to real multimodal interactive systems able to analyse gestures of many performers on stage (dancers, musicians, actors) and to map them onto real-time control of elements of the performance and/or generation of suitable audio and visual content (sound, images, lights, mobile scenery). A relevant example of exploitation of multimedia interactive systems for large scale performances is "Cronaca del Luogo" by Luciano Berio (opening of Salzburg Festival, 1999) in which the EyesWeb system (www.eyesweb.org) was used to control in real-time the processing of the voice of the main character depending on an analysis of the performer's gestures [11]. More recent examples include the public performances in the framework of the EU-IST Project MEGA (Multisensory Expressive Gesture Applications, www.megaproject.org, November 2000 – October 2003) ranging from medium-scale events, e.g., the concert "Allegoria dell'opinione verbale", by Roberto Doati, to large-scale events, e.g., "Medea" by Adriano Guarnieri [14].

The design of such multimodal interactive systems is challenging and many research issues have to be faced. For example, systems must be endowed with the capability of interpreting performers' gestures. Information contained in performers' gesture may be structured on several layers of complexity. However, in the scenario of performing arts a particular emphasis is often put on affective, expressive, emotional information. In fact, it is the capability of interpreting expressive information that allows interaction of technology and art at the level of the language art employs to convey content and to provide the audience with an aesthetical experience.

In this framework, a central role is assumed by research on *expressive gesture* [4], i.e., on the high-level emotional, affective content gesture conveys, on how to analyse and process this content, on how to use it in the development of innovative multimodal interactive systems able to provide users with natural expressive interfaces [6]. Research on expressive gesture finds its basis on previous researches on Affective Computing [21], on KANSEI Information Processing [16], on more recent studies on communication of expressive content or "implicit messages" [15], on works by psychologists (e.g., [2], [3], [17], [22], [26], [28]) and - from art and humanities - on theories from choreography (e.g., [18][19]) and music composition (e.g., [25]).

Another key issue is the development of strategies for controlling and/or generating audio and visual output in real-time. That is, even if algorithms able to correctly and reliably interpret high-level expressive information from gesture were available, the problem of if and how to use such information in an artistic performance still remains very open. The problem is even more difficult to face since it directly involves the artistic choices of the designer of the performance, i.e., how much degrees of freedom the designer wishes to leave to automatic systems: in other words the role of technology in the artwork and, from a certain point of view, the concept of artwork. These aspects have been partially faced with the definition of the concept of *expressive autonomy* [12] and have been further investigated in the framework of the MEGA project with particular reference to the definition of a conceptual architecture for modelling possible control (mapping) strategies.

This paper explores these two main issues: it focuses on gestural control of sound and visual output. After a survey of some techniques for multimodal analysis of expressive gesture developed at the InfoMus Lab of DIST - University of Genova, discussion moves on modelling of strategies for expressive control and generation of audio and visual output. Finally, some examples of applications in performing arts are given.

2. TECHNIQUES FOR MULTIMODAL ANALYSIS OF EXPRESSIVE GESTURE

Several problems have to be faced when analysing expressive gesture. Firstly, there is the need of identifying a collection of descriptors (cues) that can be used for describing expressive gesture. Secondly, algorithms have to be defined and implemented to extract measures for such descriptors. Finally, data analysis has to be performed on these measures for obtaining high-level information. Given this very rough summarization of the analysis process, this section provides a quick survey of our research on algorithms for extraction of cues at several level of abstraction, from low-level signal-related cues to high-level analysis of expressive content. Such algorithms provide the input for the strategies for gestural control of multimedia output. A discussion about two possible approaches for identifying expressive cues can be found in [8].

Expressive cues are likely to be structured on several layers of complexity. In analysis of dance fragments with videocameras, for example, some cues can be directly measured on the video frames. Others may need more elaborate processing: e.g., it may be needed to identify and separate expressive gestures in a movement sequence for computing features that are strictly related to single gestures (e.g., duration, directness).

For this reason, in the EU-IST project MEGA a conceptual framework for expressive gesture processing has been defined, structured on four layers [4][10].

Layer 1 (*Physical Signals*) includes algorithms for gathering data captured by sensors such as videocameras, microphones, on-body sensors (e.g., accelerometers), sensors of a robotic system, environmental sensors.

Layer 2 (*Low-level features*) extracts from the sensors data a collection of low-level cues describing the gesture being performed. In case of dance, for example, cues include kinematical measures (speed, acceleration of body parts), detected amount of motion, amount of body contraction/expansion. Figure 1 shows two of them.

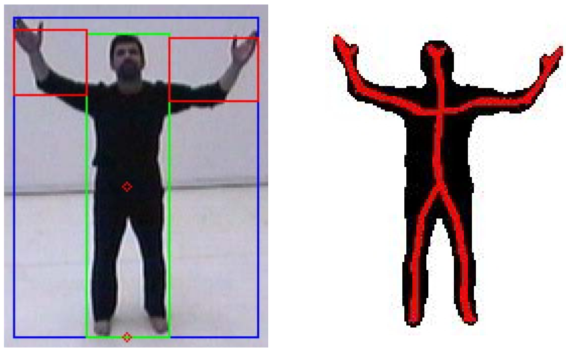
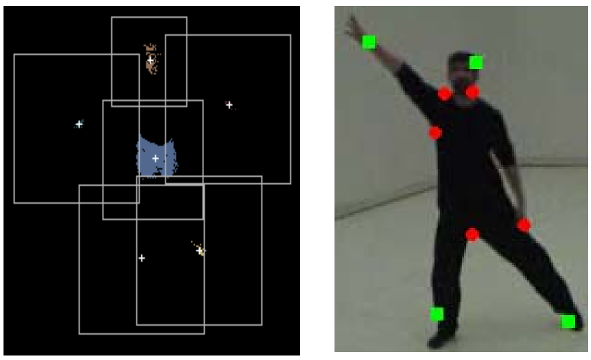


Figure 1. Examples of low-level motion cues.

The cues have been extracted from a microdance using the EyesWeb open platform (www.eyesweb.org) and, in particular, the EyesWeb Expressive Gesture Processing Library [5]. In the figure on the left some sub-regions of the body are individuated together with the body barycentre. The temporal evolution of both sub-regions and the barycentre can be analysed. In the figure on the right the skeleton of the performer extracted in real-time is represented. Some cues are currently subject of experiments aiming at (i) validating the algorithms employed for measuring them, and (ii) understanding how much these cues are really important in motion perception and in expressive content communication. For example, in a recent experiment [7] we studied the relevance of the motion of the barycentre for motion perception and for expectation in dance.

A further research direction, followed in the framework of the EU-IST project TAI-CHI (Tangible

Acoustic Interfaces for Computer Human Interaction), consists of designing systems for real-time 3D tracking of relevant body parts (i.e., usually head, hands, and feet). The images from two or more videocameras are possibly integrated with tangible acoustic interfaces for multimodal analysis. Our approach integrates several motion cues in order to obtain reliable information about the current 3D position of the tracked points. At the moment, the EyesWeb Expressive Gesture Processing Library contains several modules for 2D and 3D tracking: besides the *body skeleton* and the *sub-regions* modules sketched above, other available modules include a colour blob tracker (see Figure 2a), modules for extracting convexity defects, modules for computing limbs' extreme points based on the body contour (see Figure 2b), modules for estimating future positions of tracked points (e.g., based on Kalman filtering). A meta-layer of analysis gathers and integrates the data coming from such modules, by applying suitable rules and heuristics, to provide an approximation of the current position of the tracked points. Similarly, 2D tracking information coming from each videocamera is integrated in order to provide 3D positions of the tracked points. Even if the tracking precision is lower than in commercial dedicated motion tracking systems, the advantages (e.g., markerless system) and the obtained data are reliable enough for expressive gestural control in several multimodal interactive applications and in tangible acoustic interfaces. Moreover, the trajectories of the tracked points are further analysed by the higher layers in order to extract trajectory related expressive cues (e.g., directness, fluency, impulsiveness).



(a) (b)

Figure 2. The output of some EyesWeb modules for 2D and 3D tracking: (a) Color blob tracking; head, hands, and feet are tracked; the rectangles around each point represent the areas where the module will look for that point in the next frame. (b) Limbs' extreme points (squares) and convexity defects (circles): both are based on body contour processing. Note that in Figure 2b the left hand is labeled as a convexity defect. In fact, the left arm and hand lean on the body and the hand is the point of the body contour locally at the maximum distance from the body convex hull (i.e., it is a convexity defect).

Layer 3 (*Mid-level features and maps*) deals with two main issues in multimodal expressive gesture analysis: segmentation of the input stream (movement, music) in its composing gestures, and representation of such gestures in suitable spaces. Thus, the first problem here is to identify relevant segments in the input stream and associate to them the cues deemed important for expressive communication. For example, in dance analysis a fragment of a performance might be segmented into a sequence of gestures where gesture's boundaries are detected by studying velocity and direction variations. Measurements performed on a gesture are translated to a vector that identifies it in a semantic space representing categories of semantic features related to emotion and expression. Sequences of gestures in space and time are therefore transformed in trajectories in such a semantic space. Trajectories can then be analysed e.g., in order to find similarities among them and to group them in clusters. In Figure 3 an example of such process is shown: gestures are represented in a 2D space whose X axis represents Quantity of Motion (i.e., the amount of detected motion) while Y axis is Fluency. The analysis of the trajectories in the space was used for the real-time dynamic interpretation of two pieces of classical music: a neutral music score was dynamically interpreted and played (in a heavy, light, hard, soft way) depending if the same expressive intention was detected in the input gestures (demo in collaboration with DEI-CSC University of Padova at IBC2001, Amsterdam) [10].

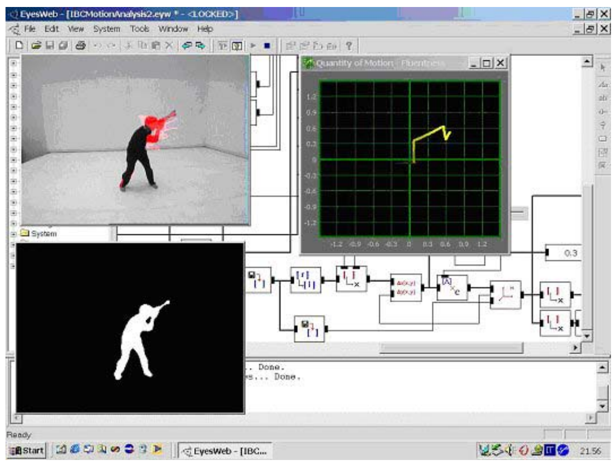


Figure 3. Gesture represented as trajectory in a 2D space (X axis: Quantity of Motion, Y axis: Fluency).

Layer 4 (*Concepts and structures*) is directly involved in data analysis and in extraction of high-level expressive information. In principle, it can be conceived as a conceptual network mapping the extracted features and gestures into (verbal) conceptual structures. For example, a dance performance can be analysed in term of the performer's conveyed emotional intentions, e.g., the basic emotions anger, fear grief, and joy. However, other outputs are also possible: for example, a structure can be envisaged describing the Laban's conceptual framework of gesture Effort, i.e., Laban's types of Effort such as "pushing", "gliding", etc. (see [18][19]). Experiments can also be carried out aiming at modelling spectators' engagement. Machine learning techniques can be employed ranging from statistical techniques (e.g., multiple regression and generalized linear techniques), to fuzzy logics or probabilistic reasoning systems (e.g., Bayesian networks), to various kinds of neural networks (e.g., classical back-propagation networks, Kohonen networks), support vector machines, decision trees. In a recent experiment described in [4] we tried to classify expressive gesture in dance performance in term of the four basic emotions anger, fear, grief, and joy. Results showed a rate of correct classification for the automatic system (five decision tree models) in between chance level and spectators' rate of correct classification. In another experiment, discussed in the same paper, we measured the engagement of listeners of a music performance (a Skriabin's Etude) and analysed correlations with extracted audio cues and with cues obtained from the movement of the performer (a pianist).

3. STRATEGIES FOR EXPRESSIVE GESTURAL CONTROL IN MULTIMODAL INTERACTIVE SYSTEMS

In multimodal interactive systems in general, and in multimodal interactive systems for performing arts in particular, control models can often be placed along a continuum having completely reactive strategies on one extreme, and real-time high-level dialogue among system and performers on the other extreme (see Figure 4). In between such two extremes several different conditions can be found with different degrees in merging reactive and dialogical behaviour. Such different conditions have a strong impact with respect to expressive autonomy (usually low for reactive behaviour and high in case of dialog), to the complexity of the automatic system (increasing with expressive autonomy) both in terms of difficulties in the design of the system and of its computational load at run-time, to temporal constraints (and in particular real-time constraints). For example, a completely reactive system will have to produce its output in a time short enough to allow perception by spectators of the association between performer's action and system's reaction. Systems based on dialog will have to produce their output in a suitable time with respect to the timing of the dialog.

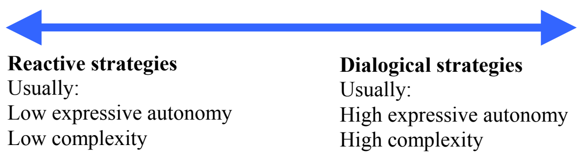


Figure 4. The control strategies continuum.

A further aspect is the possibility of employing expressive information in the control strategies. In case of dialogical strategies, for example, multimodal interactive systems able to process such information can benefit of this ability to make the dialog more effective.

A first model for control strategies (sometimes also called interaction strategies or mapping strategies) has been worked out in the framework of the MEGA project. This work led to the definition of an expressive mapping component taking into account the layered structure of the input expressive cues. The component is part of a more complex framework including also components for expressive gesture analysis and expressive gesture synthesis. All the three components are used as building blocks of an architecture for virtual subjects inhabiting expressive and distributed mixed reality performance spaces (see [27]). The expressive component (or part of it) can be employed as guideline for designing reactive systems, dialogical systems, and their combinations.

The structure of the expressive mapping component is sketched in Figure 5.

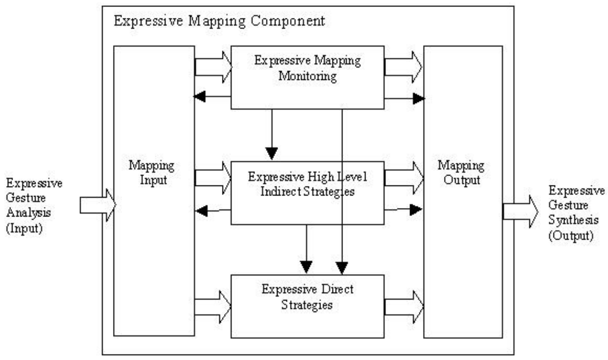


Figure 5. A model for expressive gestural control strategies structured on three processing layers. Three basic modules are considered: direct strategies allowing the implementation of reactive behaviours, high-level indirect strategies more related to rational and cognitive processes, mapping monitoring for adapting control strategies to the context and maximise mapping effectiveness. White and thick arrows represent flows of data between the components, while black and thin arrows represent influences that a component exerts on another one.

The work of the component can be illustrated with an example: consider an artistic performance employing an on-stage multimodal interactive system. Several people (dancers, musicians, actors) are on the stage. Moreover, the stage contains fixed and mobile (e.g., robot) scenery, and virtual elements (e.g., information associated to particular paths). The system observes the environment and analyses the expressive gestures of the performers. The expressive information extracted by the four layers of analysis is fed as input to the control strategies. Low-level cues produce immediate responses: e.g., an increased energy in the motion of a dancer could produce an increased rhythm in percussions; a crescendo of a pianist can produce more vivid colours in projected abstract shapes. High-level cues produce slower but continuous changes in the context of the performance: a slow transition between a rigid and angry movement toward a smooth and joyful one produces a gradual and continuous change in the association of dancer movements to sound and musical instruments, e.g., from percussions to strings.

Similarly to cue extraction, control strategies are also structured on several layers. In particular, in this model the expressive mapping component includes three layers:

- Expressive direct strategies

- Expressive high-level indirect strategies
- Expressive mapping monitoring

Expressive direct strategies consist of associations without any dynamics of expressive cues of analysed expressive gestures with parameters of synthesised expressive gestures (e.g., parameters of sound synthesis models). For example, the actual position of a dancer on the stage can be mapped onto the reproduction of a given sound. Expressive direct strategies are often associated with the lower levels of the conceptual framework discussed in the previous section: for example parameters extracted in Layer 2 (e.g., amount of motion – loudness) can be used to control particular features in the real-time generation of audio and visual content. Expressive direct strategies are usually employed for implementing reactive behaviour. Systems explicitly designed for having only a reactive behaviour (i.e., placed at the left end of the continuum in Figure 4) may consist of only expressive direct strategies. Several possible implementations are available for these strategies. One of them consists of collections of pre-defined condition-action rules, i.e., set of rules associating given configurations of parameters coming from the analysis side with given configurations of synthesis parameters. Another one employs collections of algebraic functions, computing values of synthesis parameters depending on values of analysed expressive cues. It should be noticed that while the complexity of an algebraic function can be freely increased according to any possible need, it anyway remains a static function, i.e., the mapping it induces does not change anymore once the function is defined and put at work.

Expressive high-level indirect strategies may include reasoning and decision-making processes and are often related to rational and cognitive processes. For example, consider a software module able to make decisions based on the incoming decoded expressive content: it could select an algebraic function (an expressive direct strategy) within a collection of possible algebraic functions, thus allowing direct strategies to be adapted to the current context, i.e., implementing an adaptive and dynamic direct mapping. Indirect strategies are usually characterized by:

- A state evolving over time (that is, they are dynamic processes): such a state can be updated for example by applying some kind of reasoning technique to the available information.
- Decisional processes, i.e., the system can make decisions based on the incoming information from analysis and the acquired knowledge. Such decisions may concern the kind of expressive content to produce and how to convey it, and can be related for example to the narrative structure of a performance.

Production systems and decision-making algorithms can be employed to implement this kind of strategies. As observed above, indirect strategies may also intervene on direct strategies. Performances based on the dialogical paradigm can employ indirect strategies only or a suitable mix of direct and indirect strategies.

Expressive mapping monitoring is a further layer of processing influencing both direct and indirect strategies. It consists of algorithms trying to measure the effectiveness of the lower mapping layers with respect to the overall goals of the performance. Effectiveness could be considered under several aspects: for example, in artistic performances it could be related to the audience's engagement; in a museum scenario it could be associated to visitors' fruition of the museum exhibit. Such a measure could be the result of a direct evaluation by spectators, in case it is not possible to calculate it automatically. As result of such analysis, the expressive mapping monitoring module can modify and adapt the processing of the lower mapping layers (e.g., by modifying decision parameters or changing possible collection of algebraic functions) in order to maximise effectiveness.

In the current version of the model mapping input and output modules just play the role of adapters with respect to the analysis and synthesis components (e.g., for possible conversions among different formats for data).

We plan to fully implement the model in the EyesWeb open platform through a sub-patching mechanism and the development of a further layer of processing at a higher level of abstraction (what we call *META-EyesWeb*) able to supervise and to schedule execution of patches and subpatches according to adaptive narrative structures. For example, the META-EyesWeb layer can support simple timelines of activation of patches in live electronics performances. But, more interestingly, it supports a dynamic graph of execution (i.e., an interactive narrative structure) where each node is a (sub)patch and each link defines the semantics on how to pass from its input patch to its output patch. For example, it can be defined a "fading" behaviour between two patches, whose parameters can depend on previous history and concurrent active patches.

4. EXAMPLES AND CONCLUSIONS

The strategies for gestural control discussed in this paper have been employed for artistic performances in several public events, especially in the EU-IST project MEGA. A quite recent example is an event planned in the framework of the New York University's Music and Dance Program in Italy (2003 edition), a three weeks Summer School held in Genova (director of the program Esther Lamneck). The program collected artists and multimedia experts teaching courses on arts and multimedia to a selected number of students. In the 2003 edition, the event consisted of an installation and a concert. The concert (lasted for about one hour) consisted of six pieces: "Madrid" (Full ensemble and tape), by Keith Fullerton Whitman, "Scrivo in Vento" (Interactive Performance) by Elliot Carter, "WZJB" (Full ensemble and tape) by William Raynovich, "On edge" (Full ensemble) piece composed and prepared during the NYU Summer School in Genova, “Tarogato” by Larry Austin and Esther Lamneck, “Mappaemundi” by Lawrence Fritts, and “Rock and Roll Goddess” by Eric Lyon. All these pieces included also dance performance (choreographer Douglas Dunn).

Four of the six pieces included real-time interaction with EyesWeb and employed different strategies for real-time gestural control of audio and visual output. Both musicians and dancers could interact with the system performing real-time analysis of both audio and movement. Control strategies were developed during the summer school and refined in the rehearsals. For example, Figure 6 shows a piece in which both reactive and dialogical behaviour was implemented. Visual output (abstract shapes obtained from the silhouettes of the dancers moving along the edge of the stage) directly depended on analysed features of the music performers' expressive gestures, while at the same time a dialog was created among the played music and real-time computer synthesised sound.

Figure 6. Excerpt from the NYU concert in Genova, Auditorium “E. Montale”, Opera House Carlo Felice, Genova, Italy, July 19th, 2003. Director Esther Lamneck, choreographer Douglas Dunn, musicians and dancers the NYU students, interactive systems by DIST-InfoMus Lab and Eidomedia, photo by Matteo Ricchetti.

Art, however, is not the only application field that can benefit of research on multimodal interactive systems. Another domain of interest is therapy and rehabilitation in which we carried out some pilot experiments. In the framework of the EU-IST project CARE HERE, for example, prototypes of multimodal interactive systems have been developed to analyse body movements of different kinds of patients (Parkinson's patients, severely handicapped children, people with disabilities in the learning processes) and to map the analysed parameters onto automatic real-time generation of audio and visual outputs, attempting to create aesthetic resonance [9].

A particular focus of our current research is on Tangible Acoustic Interfaces (TAI) that employ physical objects and space as media to bridge the gap between virtual and physical worlds and to make information accessible through touchable objects as well as through ambient media. TAI are addressed in the framework of the EU-IST project TAI-CHI (2004-2006), whose primary objective is the development of acoustic-based remote sensing technologies which can be adapted to virtually any physical object to create tangible interfaces, allowing the user to communicate freely and naturally with a computer, an interactive system, or the cyber-world.

Other projects are related to the development of enhanced interfaces for human-computer interaction (the EU-IST Network of Excellence ENACTIVE), to the investigation of the role of emotion in non-verbal communication (the the EU-IST Network of Excellence HUMAINE), to sound perception and production (the EU-IST Coordinated Action S2S^2).

5. ACKNOWLEDGEMENTS

We thank our colleagues at the InfoMus Lab who contributed to this research with useful discussions. In particular we thank Paolo Coletta, Alberto Massari, Massimiliano Peri, Matteo Ricchetti, and Andrea Ricci. We also thank the Director of the New York University Music and Dance Program in Italy Prof. Esther Lamneck, the choreographer Douglas Dunn, and the NYU students. Some parts of the research discussed in this paper have been partially supported by the EU-IST project TAI-CHI.

6. REFERENCES

[1] Arfib, D., Couturier, J. M., Kessous, L. “Design and use of some new digital musical instruments”, in *A. Camurri, G. Volpe (Eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915*, Springer Verlag, Heidelberg, Germany, 2004.

[2] Argyle, M. *Bodily Communication*. Methuen & Co Ltd, London, UK, 1980.

[3] Boone, R. T., Cunningham, J. G. “Children's decoding of emotion in expressive body movement: The development of cue attunement”, *Developmental Psychology*, 34: 1007-1016, 1998.

[4] Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., Volpe, G. “Multimodal analysis of expressive gesture in music and dance performances”, in *A. Camurri, G. Volpe (Eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915*, Springer Verlag, Heidelberg, Germany, 2004.

[5] Camurri, A., Mazzarino, B., Volpe, G. “Analysis of Expressive Gesture: The EyesWeb Expressive Gesture Processing Library”, in *A. Camurri, G. Volpe (Eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915*, Springer Verlag, Heidelberg, Germany, 2004.

[6] Camurri, A., Mazzarino, B., Volpe, G. “Expressive interfaces”, *Cognition,*

Technology & Work, 6(1): 15-22, Springer-Verlag, London, UK, 2004.

[7] Camurri, A., Krumhansl, C. L., Mazzarino, B, Volpe, G. "An Exploratory Study of Anticipating Human Movement in Dance", in *Proceedings 2nd International Symposium on Measurement, Analysis and Modeling of Human Functions*, Genova, Italy, 2004.

[8] Camurri, A., Mazzarino, B., Menocci, S., Rocca, E., Vallone, I., Volpe, G. "Expressive gesture and multimodal interactive systems", in *Proceedings AISB 2004 Convention: Motion, Emotion and Cognition*, Leeds, UK, 2004.

[9] Camurri, A., Mazzarino, B., Volpe, G., Morasso, P., Priano, F., Re, C., "Application of multimedia techniques in the physical rehabilitation of Parkinson's patients", *Journal of Visualization and Computer Animation*, 14(5): 269-278, John Wiley & Sons Ltd., Chichester, UK, 2003.

[10] Camurri, A., De Poli, G., Leman, M., Volpe, G. "Toward Communicating Expressiveness and Affect in Multimodal Interactive Systems for Performing Art and Cultural Applications", *IEEE Multimedia Magazine*, IEEE CS Press, in press.

[11] Camurri, A., Hashimoto, S., Ricchetti, M., Trocca, R., Suzuki, K., Volpe, G. "EyesWeb – Toward Gesture and Affect Recognition in Interactive Dance and Music Systems" *Computer Music Journal*, 24(1): 57-69, MIT Press, Cambridge (MA), USA, 2000.

[12] Camurri, A., Coletta, P., Ricchetti, M., Volpe, G., "Expressiveness and Physicality in Interaction", *Journal of New Music Research*, 29(3): 187-198, Swets & Zeitlinger, Lisse, The Netherlands, 2000.

[13] Camurri, A., Ferentino, P. "Interactive environments for music and multimedia", *Multimedia Systems* 7: 32 – 47, Springer-Verlag, 1999.

[14] de Götzen, A. "Enhancing Engagement in Multimodality Environments by Sound Movements in a Virtual Space", *IEEE Multimedia Magazine*, April – June 2004, 4-7 IEEE CS Press, 2004.

[15] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J. "Emotion Recognition in Human-Computer Interaction", *IEEE Signal Processing Magazine,* 1, IEEE CS Press, 2001.

[16] Hashimoto, S., "KANSEI as the Third Target of Information Processing and Related Topics in Japan", in *Camurri A. (Ed.) "Proceedings of the International Workshop on KANSEI: The technology of emotion"*, AIMI (Italian Computer Music Association) and DIST-University of Genova, 101-104, Genova, Italy, 1997.

[17] Krumhansl, C. L., "Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No. 15", *Musicae Scientiae*, 1, 63-85, 1997.

[18] Laban, R., Lawrence, F.C. *Effort*. Macdonald & Evans Ltd., London, UK, 1947.

[19] Laban, R. *Modern Educational Dance*. Macdonald & Evans Ltd., London, UK. 1963.

[20] Machover, T., Chung, J. "Hyperinstruments: Musically intelligent and interactive performance and creativity systems", in *Proceedings International Computer Music Conference 1989 (ICMC89)*, 186-190, 1989.

[21] Picard, R. *Affective Computing*. Cambridge, MA, MIT Press, 1997.

[22] Pollick, F. E., "The Features People Use to Recognize Human Movement Style", in *A. Camurri, G. Volpe (Eds.), Gesture-based Communication in Human-Computer Interaction, LNAI 2915*, Springer Verlag, Heidelberg, Germany, 2004.

[23] Rowe, R. *Machine Musicianship*. Cambridge, MA, MIT Press, 2001.

[24] Rowe, R. *Interactive music systems: Machine listening and composition*. Cambridge, MA, MIT Press, 1993.

[25] Schaeffer, P. *Traité des Objets Musicaux.* Second Edition, Editions du Seuil, Paris, France, 1977.

[26] Scherer, K.R. "Why music does not produce basic emotions: pleading for a new approach to measuring the emotional effects of music", in *Proceedings Stockholm Music Acoustics Conference SMAC-03*, 25-28, KTH, Stockholm, Sweden, 2003.

[27] Volpe, G. *Computational models of expressive gesture in multimedia systems.* Ph.D. Dissertation, Faculty of Engineering, University of Genova, Genova, Italy, 2003.

[28] Wallbott, H.G. "The measurement of Human Expressions", in *Walbunga von Rallfer-Engel (Ed.) Aspects of communications*, 203-228, 1980.