Non-verbal Full Body Emotional and Social Interaction: A Case Study on Multimedia Systems for Active Music Listening

Antonio Camurri

InfoMus Lab – Casa Paganini Intl Research Centre
DIST- University of Genova
Piazza Santa Maria in Passione 34, 16123 Genova, Italy
antonio.camurri@unige.it

Abstract. Research on HCI and multimedia systems for art and entertainment based on non-verbal, full-body, emotional and social interaction is the main topic of this paper. A short review of previous research projects in this area at our centre are presented, to introduce the main issues discussed in the paper. In particular, a case study based on novel paradigms of social active music listening is presented. Active music listening experience enables users to dynamically mould expressive performance of music and of audiovisual content. This research is partially supported by the 7FP EU-ICT Project SAME (Sound and Music for Everyone, Everyday, Everywhere, Every Way, www.sameproject.eu).

Keywords: non-verbal full-body multimodal interfaces, emotion, social signals, sound and music computing, active listening.

1 Introduction

Entertainment applications characterized by full-body engagement of users has emerged in the recent years. The Nintendo Wii and the diffusion and familiarity by users with interactive art multimedia installations (e.g. in museums and science centers, in public spaces) are examples that bear witness of the diffusion of such "embodied" applications. Art and entertainment is contributing to move toward more effective, user-centric, embodied paradigms of interaction with multimedia content. In a broader scenario, we can envisage future internet networked media enabling users within the next 5-10 years to be fully immersed with 3D audiovisual content, using natural interfaces sensitive to the context and to full-body multimodal expressive intentions, and with the highest degrees of physical as well as social engagement. See for example the User Centric Media EU FP7 ICT projects for a state of the art on these visions.

Already in the sixties, the choreographer Merce Cunningham and the composer John Cage developed an artistic project where a dancer could move in a sensitive space, where his body could modify electronically generated music: this was a sort of a large scale Theremin musical instrument, and we can consider it as a pioneer work

A. Nijholt, D. Reidsma, and H. Hondorp (Eds.): INTETAIN 2009, LNICST 9, pp. 9–18, 2009.

anticipating current trends of interactive entertainment, interactive dance and music systems, and interactive installations. Many attempts have been done by researchers, artists, and practitioners until nowadays in this direction. International conferences such as ISEA (International Symposium on Electronic Arts), NIME (New Interfaces for Musical Expression, www.nime.org), and ICMC (International Computer Music Conference) bear witness of these developments along more than three decades. A number of challenging and significant examples of cross-fertilization between artistic and scientific/technological research emerged: one example among the several available, can be found in "digital luthery", where the longstanding experience of traditional lathery, developed in many centuries of evolution of musical instruments, is exploited for enhancing and developing not only novel digital musical instruments, but also novel human-computer interfaces and interaction paradigms.

We focus here on innovative multimedia and multimodal human-computer interfaces and systems for performing arts, museums, audience and spectators interfaces, systems for therapy and rehabilitation (we may call it "entertherapy"): these systems are characterized by non-verbal, full-body, emotional, and social interaction. We aim at understanding, modeling, and exploiting the non-verbal expressive gestures (e.g. [1]) and social signals (e.g. [2]), to develop innovative intelligent interfaces and multimedia systems. Our approach is based on the cross-fertilization of scientific and artistic research. Artistic productions [3, 4], museum and science-centre applications [5], therapy and rehabilitation ("entertherapy") [6], and music and multimedia industry entertainment [7, 8] are the main targets where our research results are applied and validated. At the same time, artistic and humanistic theories are often sources of inspiration for scientific research. For example, the Theory of Effort by the choreographer Rudolf Laban has been utilised to design computational models of expressive gesture. In another occasion, important inspiration to scientific research came from the Morphology theory by the French composer Pierre Schaeffer (e.g. [1, 4]).

Given these premises, we briefly sketch in Section 2 a few examples from our research, started in the eighties, toward interactive systems for performing arts and entertainment, based on non-verbal, multimodal, expressive and emotional interaction. Then, the paper considers a main case study on social active music listening, one of our current research project at Casa Paganini – InfoMus Lab. Novel paradigms for social active music experience, based on the dynamic moulding of audiovisual content and expressive music performance are described. This research finds its origins from previous projects (the EU Esprit Project MIAMI – Multimodal Interaction for Advanced Multimedia Interfaces, 1995-1997; the EU CRAFT Project U-CREATE: the EU IST Project MEGA – Multisensory Expressive Gesture Applications, www.megaproject.org), and it is the core topic of the recently started 7FP EU-ICT Project SAME (Sound and Music for Everyone, Everyday, Everywhere, Every Way, www.sameproject.eu). Further, the case study has been designed and developed using the EyesWeb XMI open software platform (www.eyesweb.org).

2 Previous Research and Systems

Since the late eighties, research at InfoMus Lab (www.infomus.org) was directed to the study of sound and music computing, focusing on the understanding of sonic and music communication. This was since the beginning grounded on the hypothesis that knowledge about sound and music must include models on both the motion alluded, evoked by music and the real motor behavior needed to produce the music: music originates by movement, music evokes movement (e.g. in dance), the gestural component is fundamental to understand sonic and music content, up to the expressive subtleties typical of music interpretation in a performance [13].

The HARP approach and system [7] originated a number of entertainment and artistic applications in which music content is related to motor activity. Figure 1 shows the conceptual model of the HARP system. In [7] a series of interactive applications emerging from the research on HARP were proposed (see figure 2): the SoundCage system was based on a sort of sensorized cage equipped with about 80 IR sensors and floor tiles (see figure 2, on the left), capable to reconstruct approximated features of the movement of a human moving inside the "cage"; the HARP/Vscope was based on the V-Scope sensor system, characterized by IR / US wearable wireless markers able to track human behavior; the HARP/DanceWeb was based on a cluster of US sensors directed vertically from the ceiling to the floor, integrated with sensorized pressure tiles. Description of the latter two systems are available in [5] and in [3]. Main ideas were to influence and modify the music listened by dancing, and to develop novel concepts of games, we called them "audiogames", where sonic and music material were the main content, addressed and manipulated by the non-verbal full-body behavior, keeping into account memory and context. HARP was a hybrid system to support the development of interactive gesture- and sound-driven interactive stories. We proposed several proptotypes and museum applications based on HARP, for example an application able to re-mix and re-structure in real-time MIDI-based music

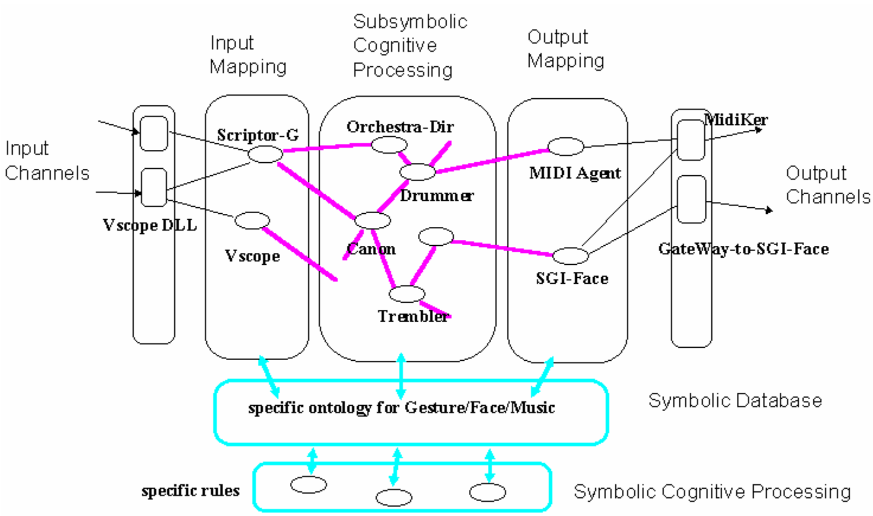


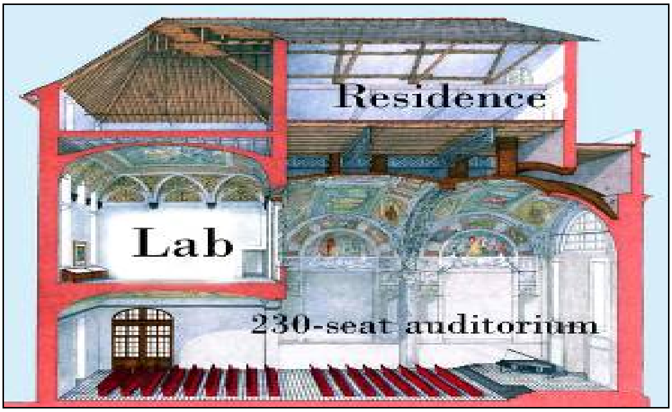
Fig. 1. The HARP model for an application where human gesture controls both the synthesis of the emotion of an artificially synthesized 3D human face and the interpretation of a computer-generated music [7]. This was designed in the MIAMI EU Esprit project (1994-1997).

Fig. 2. Three interactive dance-music systems: the HARP/DanceWeb (in the picture, it is the "start"-like structure hanging on the right from the ceiling), the HARP/Vscope (in the centre, on the floor the receivers of the signals from wireless markers are shown), and the SoundCage interactive dance-music machine (the sensorized "cage" on the left and behind the column), at the EC EITC'95 Exhibition as entertainment tools in the CyberCafé' area (Brussels Congress Centre, 27-29 Nov.1995.

pieces, thus obtaining a sort of "dance karaoke", where orchestration, tuning, quality of the sound depends on the quality of dance. The diverse scenarios of entertainment applications can be found in [3, 7].

The idea of active experience of audiovisual content was also the basic idea for the museum project described in [5]: the "Music Atelier" of the "Città dei Bambini" science center in Genoa included a series of interactive installations: one was based on the HARP/DanceWeb for interactive audio-based storytelling, to explore and experience physically sound and music content. Further, a simple mobile robot, based on Stanford Research Institute "Pioneer 1" platform, was acting in the "Music Atelier" as a mobile installation, a companion of children, talking and interacting with them and the other installation of the atelier. The Music Atelier was active for five years from 1997, and was one of the first projects of our Centre for science-centres and museums.

From 1997 to 2004 InfoMus Lab was hosted at the Genoa Opera House - Teatro Carlo Felice: this collaboration included a lab site at the theatre and joint activities, including projects involving scientific experiments within artistic productions of the Opera House. This was very productive for maturing experience on scientific and technological research development cycle, grounded on the cross-fertilisation of artistic and scientific research. For example, we had the opportunity to participate to



(a)

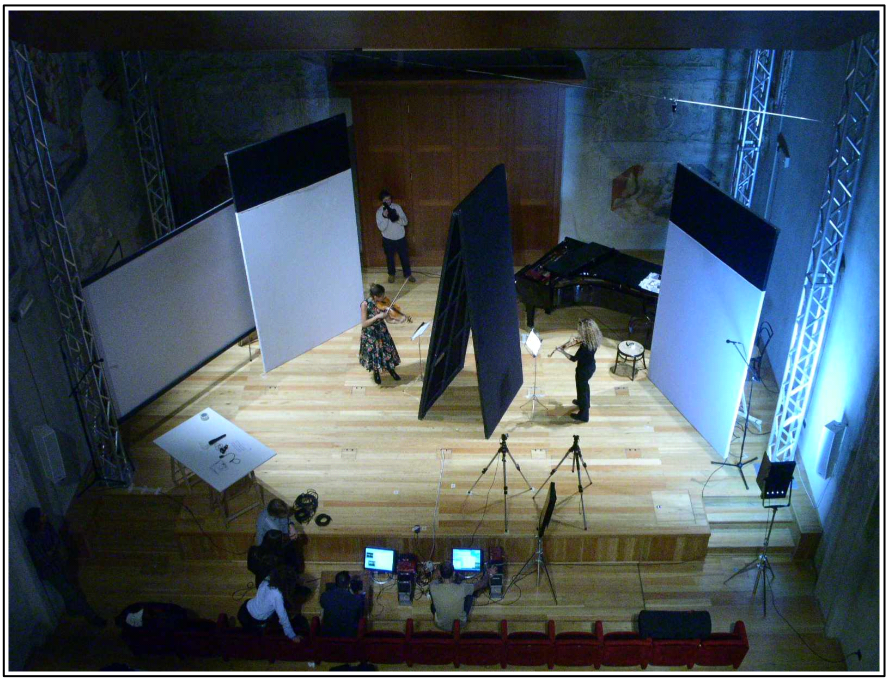


Fig. 3. (a) the layout of two of the rooms of *Casa Paganini – InfoMus Intl Centre of Excellence*: one of the main Lab rooms (Matroneo) is in a space not visible from the auditorium but allowing a full control of what happens on stage: this allows researchers to perform ecological experiments on emotion and empathy (among musicians, dancers, or between performers and the audience) without any perturbation. (b) the experiment in 2006, in occasion of the *Intl Violin Competition "Premio Paganini"* and the *EU Summer School of the Network of Excellence HUMAINE*: selected violinists were measured in different conditions (eye Vs no-eye contact; induced Vs simulated emotion; etc.), in order to create a dataset to study emotions and in particular emotional entrainment between musicians [12], and with the audience during a concert. The view of the stage is as it is seen from the Lab/Matroneo.

the artistic project of the "lighting keyboard" for the Scriabin's "Promethee" symphonic poem, performed with the orchestra of the Opera House. Here the interaction was between the colors evoked by the "light keyboard", one of the instruments of the orchestra (for more details, see www.infomus.org), and its relations with the music and the concert hall of the opera house, providing interesting insights on interactive systems for multiple users (the audience of the concert) integrating sound and color.

Since 2005, InfoMus Lab has its new premises in the recently restored monumental building of S.Maria delle Grazie La Nuova - Casa Paganini. This new site is determinant to contribute to the growth of the process of cross-fertilisation between artistic and scientific research (see figure 3).

3 Case Study: Social Active Music Listening

We recently started a project on active music listening (www.sameproject.org; [8]). Music making and listening are a clear example of interactive and social activity. However, nowadays mediated music making and listening is usually still a passive, non–interactive, and non-context sensitive experience. The current electronic technologies, with all their potential for interactivity and communication, have not yet been able to support and promote this essential aspect of music making and listening. This can be considered a degradation of traditional listening experience, in which the public can interact in many ways with performers to modify the expressive features of a piece.

The need of recovering such active attitude with respect to music is strongly emerging, and novel paradigms of active experience are going to be developed. With *active experience* and *active listening* we mean that listeners are enabled to interactively operate on music content, by modifying and molding it in real-time while listening. Active listening is the basic concept for a novel generation of interactive music systems [9], which are particularly addressed to a public of beginners, naïve and inexperienced users, rather than to professional musicians and composers.

Active listening is also a major focus for the EU-ICT Project SAME (Sound and Music for Everyone, Everyday, Everywhere, Every Way, www.sameproject.eu). SAME aims at: (i) defining and developing an innovative networked end-to-end research platform for novel mobile music applications, allowing new forms of participative, experience-centric, context-aware, social, shared, active listening of music; (ii) investigating and implementing novel paradigms for natural, expressive/emotional multimodal interfaces, empowering the user to influence, interact, mould, and shape the music content, by intervening actively and physically into the experience; and (iii) developing new mobile context-aware music applications, starting from the active listening paradigm, which will bring back the social and interactive aspects of music to our information technology age.

The Orchestra Explorer [8] allows users to physically navigate inside a virtual orchestra, to actively explore the music piece the orchestra is playing, to modify and mold in real-time the music performance through expressive full-body movement and gesture. Concretely, the virtual orchestra is spread over a physical surface (see figure 4). By walking and moving on the surface, the user discovers each single instrument or section and can operate through her expressive gestures on the music piece the instrument is playing.

The Orchestra Explorer is not a simple reproduction or remixing of multiple audio tracks, nor a full automatic conducting system. It is something in between these two extremes. On the one hand, it provides the active listener with means for operating on the sound and music content that are not available in simple (passive) reproduction or remixing of multiple audio tracks. On the other hand, it does not provide the full control of the performance as in traditional conducting systems (e.g., see Lee et al., 2006). This approach is motivated by our aim of developing a new paradigm which, while actively engaging listeners, is at the same time different from traditional metaphors such as conducting, and enabled by recent research on expressive multimodal interfaces. The user really becomes an explorer of sound and music, i.e., she discovers the content step by step; she gradually understands how music performance works; she learns how to operate on the content. Figure 4 shows a sketch of the Orchestra Explorer paradigm.

We developed another active music listening paradigm, starting from the *Orchestra Explorer*, characterized by social interaction among users. We applied this paradigm for the design of a multimedia installation and a performance for dancers and singers: *Mappe per Affetti Erranti* [11]. It extends and enhances the *Orchestra Explorer* in two major directions. On the one hand it reworks and extends the concept

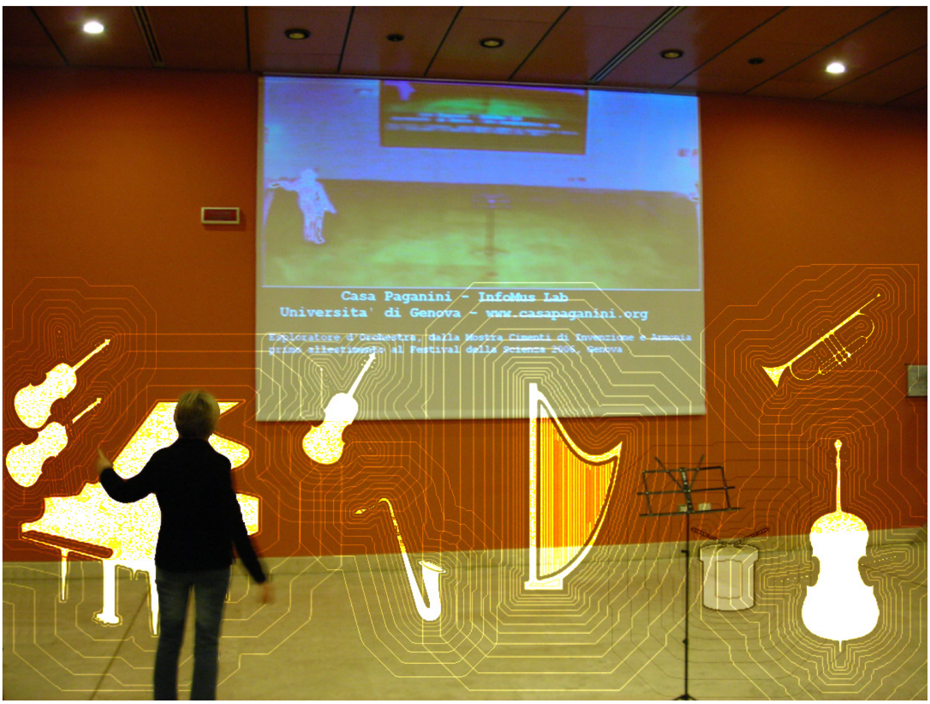


Fig. 4. An example of implementation of the paradigm of the active music listening paradigm: the "orchestra explorer" at Accademia di Santa Cecilia in Rome (2007). (Photo by Corrado Canepa)

Fig. 5. The paradigm of active music listening with social interaction: each user "embodies" a music section of the piece, and her behavior influences the expressive interpretation of her associated music section. We adopted vocal music pieces in our experiments: so each of the four users "embodies" one of the specific voices of the music piece (baritone, tenore, contralto, and soprano). The music is pre-recorded. The users can intervene to modify the voices with various degrees of freedom. For example, if a user moves shy and hesitant, her associated voice changes, e.g. toward a more intimate and whispering interpretation. If she moves aggressive, her voice performance becomes loud, extrovert, and angry. Emotion models from experimental psychology are here utilized. To obtain a coherent performance, all the users must converge to behave with similar, coherent behavior. Otherwise, the different voices remain unsynchronized and with separate, contrasting interpretations.

of navigation of a virtual orchestra by introducing multiple levels: from the navigation in a physical space populated by virtual objects or subjects (as it is for the musical instruments in the *Orchestra Explorer*) up to the navigation in virtual affective, emotional spaces populated by different expressive performances of the same music piece. Users can navigate such affective spaces by their expressive movement and gesture. On the other hand, *Mappe per Affetti Erranti* explicitly addresses the fruition by multiple users and encourages collaborative behavior: only collaboration among users brings toward aesthetically pleasant reconstructions of the music piece. In other words, while users explore the physical space, the (expressive) way in which they move and the degree of collaboration between them allow them to explore at the same time an affective, emotional space. On the one hand, the new system exploits a refined set of features for expressive gesture classification, including a new model for measurement of impulsiveness and features for analysis of the behavior of a group of users as a whole. On the other hand, it supports adaptation to context, i.e., it can change inputs, outputs, analysis and mapping strategies according to what is needed in the different situations For example, during a performance the interaction rules may change according to the evolution of the music piece. In a situation of shared/social active listening, the situation can adapt to the type and behavior of users.

The case studies described in this section (as well as all the research prototypes and applications we developed in the last ten years) have been developed with the EyesWeb open software platform. The latest version, EyesWeb XMI, is available for download from www.eyesweb.org.

4 Discussion

The social active listening paradigm described in the previous section opens new perspectives for both research and applications. We studied and developed a number of techniques for measuring social signals, resulting in a new library for our EyesWeb XMI open software platform (www.eyesweb.org). A novel approach and novel algorithms for the analysis of social behavior, and in particular of human emotional synchronization has been developed [12]. In a recent experiment, gestures from a duo of violin players were collected (see figure 3b) and analysed to detect occurrences of emotional-based Phase Synchronisation. The analysis of gesture is conducted using Recurrence Plot and Recurrence Quantification Analysis methods, and a statistical validation of results is done exploiting a surrogate data approach. For example, in this experiment, some of the results show that emotional entrainment occur preferably when players can communicate only by sound, and that positive emotions such as Serenity and Joy seem to promote this coordination better than negative emotions like, for example, Anger and Sadness.

Current research work is directed toward the refinement and extension of social descriptors for empathy, emotional entrainment, leadership, and attention. Two papers in preparation describe the results available so far.

We deem that the inclusion of social signals in the design of multimodal interfaces and systems such as the examples presented in this paper is important to move toward a novel generation of user-centric, embodied artistic and entertainment applications characterized by non-verbal, full-body interaction.

Acknowledgements

I am deeply grateful to my colleagues and friends Corrado Canepa, Paolo Coletta, Gualtiero Volpe, for their invaluable contribute to the development of the active music listening project. I wish to thank also the other colleagues and friends at InfoMus Lab, and in particular Nicola Ferrari, Donald Glowinski, Alberto Massari, Barbara Mazzarino, Maurizio Mancini, Giovanna Varni.

This research is partially supported by the 7FP EU-ICT Project SAME (Sound and Music for Everyone, Everyday, Everywhere, Every Way, www.sameproject.eu). SAME is coordinated by Casa Paganini – InfoMus, and the partners are Nokia Research Centre, Ircam, KTH – Stockholm, TKK – Helsinki Institute of Technology.

References

1. Camurri, A., Mazzarino, B., Volpe, G.: Expressive Interfaces. In: Marti, P. (ed.) Cognition Technology & Work, special issue on "Presence: design and technology challenges for cooperative activities in virtual or remote environments", vol. 6, pp. 15–22. Springer, Heidelberg (2004)
2. Pentland, A.: Socially aware, Computation and Communication. Computer. IEEE CS Press, Los Alamitos (2005)
3. Camurri, A., Ferrentino, P.: Interactive Environments for Music and Multimedia. ACM Multimedia Systems, Special issue on Audio and Multimedia 7(1), 32–47 (1999)
4. 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 12(1), 43–53 (2005)
5. Camurri, A., Coglio, A.: An Architecture for Emotional Agents. IEEE Multimedia 5(4), 24–33 (1998)
6. 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 (2003)
7. Camurri, A.: Interactive Dance/Music Systems. In: Proc. Intl. Computer Music Conference ICMC 1995, The Banff Centre for the arts, Canada, ICMA-Intl. Comp. Mus. Association, September 3-7, pp. 245–252 (1995)
8. Camurri, A., Canepa, C., Volpe, G.: Active listening to a virtual orchestra through an expressive gestural interface: The Orchestra Explorer. In: Proc. Intl Conf. NIME 2007 New Interfaces for Music Expression, New York University (2007)
9. Rowe, R.: Interactive music systems: Machine listening and composition. MIT Press, Cambridge (1993)
10. Lee, E., Karrer, T., Borchers, J.: Toward a Framework for Interactive Systems to Conduct Digital Audio and Video Streams. Computer Music Journal 30(1), 21–36 (2006)
11. Camurri, A., Canepa, C., Coletta, P., Mazzarino, B., Volpe, G.: Mappe per Affetti Erranti: a Multimodal System for Social Active Listening and Expressive Performance. In: Proc. Intl Conf. NIME 2008 – New Interfaces for Musical Expression, Casa Paganini, University of Genoa (2008), http://www.nime.org
12. Varni, G., Camurri, A., Coletta, P., Volpe, G.: Emotional Entrainment in Music Performance. In: Proc. 8th IEEE Intl Conf. on Automatic Face and Gesture Recognition, Amsterdam, September 17-19 (2008)
13. Camurri, A., Morasso, P., Tagliasco, V., Zaccaria, R.: Dance and Movement Notation. In: Morasso, P., Tagliasco, V. (eds.) Human Movement Understanding, pp. 85–124. North Holland, Amsterdam (1986)