User-Centric Context-Aware Mobile Applications for Embodied Music Listening

Antonio Camurri[1,*], Gualtiero Volpe[1], Hugues Vinet[2], Roberto Bresin[3], Marco Fabiani[3], Gaël Dubus[3], Esteban Maestre[4], Jordi Llop[4], Jari Kleimola[5], Sami Oksanen[5], Vesa Välimäki[5], and Jarno Seppanen[6]

[1] Casa Paganini - InfoMus Lab, DIST - University of Genova, Genova, Italy
antonio.camurri@unige.it
www.sameproject.eu
[2] IRCAM, Paris, France
[3] KTH School of Computer Science and Communication, Stockholm, Sweden
[4] Music Technology Group, UPF - Universitat Pompeu Fabra, Barcelona, Spain
[5] TKK, Department of Signal Processing and Acoustics, Espoo, Finland
[6] Nokia Research Center, Helsinki, Finland
Viale Causa 13, I-16145 Genova, Italy

Abstract. This paper surveys a collection of sample applications for networked user-centric context-aware embodied music listening. The applications have been designed and developed in the framework of the EU-ICT Project SAME (www.sameproject.eu) and have been presented at Agora Festival (IRCAM, Paris, France) in June 2009. All of them address in different ways the concept of embodied, active listening to music, i.e., enabling listeners to interactively operate in real-time on the music content by means of their movements and gestures as captured by mobile devices. In the occasion of the Agora Festival the applications have also been evaluated by both expert and non-expert users.

1 Introduction

The concept of User-Centric Media entails the development of new technologies enabling an active, participative, personalized experience of media. Such technologies include, for example, innovative and intelligent real-time content processing techniques, new paradigms for natural multimodal interfaces, new devices, context-awareness. Moreover, since the strong emphasis on the user, technologies for User-Centric Media cannot avoid to take into account two major aspects of human interaction and communication: embodiment and the social dimension.

In this framework, music making and listening are an excellent test-bed for technologies for future User-Centric Media, since they are a clear example of human activities that are above all interactive and social.

The EU-ICT Project SAME (Sound And Music for Everyone Everyday Everywhere Every way, www.sameproject.eu), started in January 2008 and that recently reached half of its way, aims at developing mobile context-aware music

* Corresponding author.

P. Daras and O. Mayora (Eds.): UCMedia 2009, LNICST 40, pp. 21–30, 2010.

applications for active, embodied experience of music in cooperative social environments. The project is based on the concept of *active listening*, i.e., listeners are enabled to interactively operate on (pre-recorded) music content through their movement and gesture, by modifying and molding it in real-time while listening. This is obtained through the development of a networked end-to-end platform for mobile music applications enabling 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.

Active listening is the basic concept for a novel generation of interactive music applications, particularly addressed to a general public of beginners, naïve and inexperienced users, rather than to professional musicians. A particularly relevant aspect of active listening is its social, collaborative implication: active listening enables a social, collaborative, and context aware experience of music, allowing listeners to cooperate in the real-time manipulation and re-creation of music content.

Examples of the active listening paradigm are emerging. The *Orchestra Explorer* [1] enables users to explore a space populated by virtual instruments. *Mappe per Affetti Erranti* [2] introduces multiple levels of navigation: from navigation in a physical space up to emotional spaces populated by different expressive performances of the same music piece. Users can navigate such spaces by their expressive movement and gesture. *Mappe per Affetti Erranti* also addresses experience by multiple users encouraging social behavior. The virtual air guitars [3, 4] are examples of gesture-based mobile musical instruments. They can be made easier to play than conventional musical instruments, because user's gestures can be interpreted by the computer to produce the desired output sound.

This paper surveys a first set of such mobile context-aware music applications, presented by the SAME partners at the Agora Festival (IRCAM, Paris, France, June 2009) and representing the mid-term milestone of the project (Section 2). Prototype applications were evaluated by both expert and non-expert users visiting the festival. Results from such evaluation are also discussed (Section 3).

2 Sample Applications

In the following the sample applications presented at the AGORA Festival are shortly described. Applications can be grouped depending on how they address the concept of active listening: some of them implement active listening as an exploration of the music content, others put a particular focus on the possibility of molding the expressivity of a music piece, others adopt a game-like paradigm.

Three applications are built around the concept of exploration of a pre-recorded music piece by user's movement and gesture as captured by a mobile device. The *Audio Explorer*[1] is a mobile active-listening application allowing users to interactively de-mix commercial stereo recordings into different channels while being streamed to their mobile devices, also offering interactive re-mixing possibilities based on previously separated channels. Audio separation is carried out in a server by remotely exploring the panning position of different sources (or channels) in a stereo-mastered track. Separation parameters are controlled by means of either keypad buttons or processed

[1] Contributors: Esteban Maestre, Jordi Llop, Vassilis Pantazis – UPF; Alberto Massari – DIST.

accelerometer data gathered from the mobile phone device. The separation parameters are stored in the server, so that they are shared among users, who can access to them for using them in a re-mix context: users manipulate (gain by means of either keypad buttons or processed accelerometer data) the gain and panning position of each previously separated channel within the original recording, leading to an active listening experience. An overview of the system architecture is depicted in Figure 1. Original audio tracks reside on a database in the server side and are retrieved by the user. The EyesWeb XMI platform (v. 5.0.3.0) [5] is running on the server machine, giving support for audio streaming to the mobile device, audio processing through an extended and improved VST implementation of the audio separation algorithm described in [6], and application control protocol based on Open Sound Control (OSC). The mobile phone device (Nokia N85) runs the application control interface and offers visual feedback to the user through a user-friendly GUI. The processed audio stream is received from the server and played back locally. Real-time control of separation/remixing is performed remotely from the mobile device, which is in charge of gathering and processing accelerometer data, processing key pressing for retrieval of audio and separation preset files, and display of visual feedback to the user.

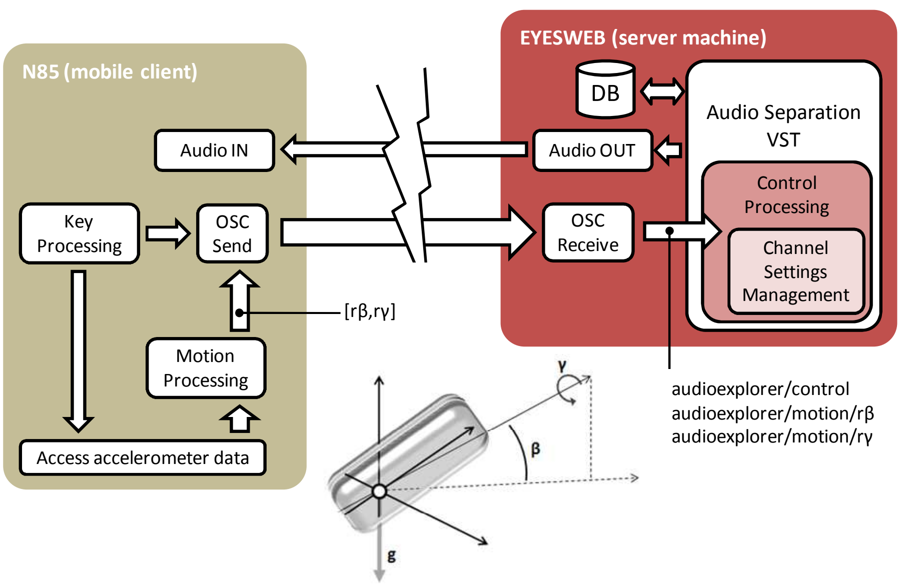

Fig. 1. System architecture of the *Audio Explorer* application

The *Mobile Orchestra Explorer*[2] is a mobile version of the former Orchestra Explorer [1]. Users can navigate a shared (physical or virtual) "orchestra space", populated by the sections or single instruments of a pre-recorded music (see Figure 2): a user can activate and listen to one or more sections of the music. The mobile phones

[2] Contributors: Antonio Camurri, Corrado Canepa, Paolo Coletta, Gualtiero Volpe, Alberto Massari, Maurizio Mancini – DIST; Markus Noisternig, Joseph Sanson, Olivier Warusfel – IRCAM for WFS extension.

are here used to detect the movement of the user, to activate and control the music sections, and to present, on the phone display, the user's position in the orchestra space. The music rendering is either based on 3D sound via loudspeakers (using WFS) or on the mobile phone using its headphones.

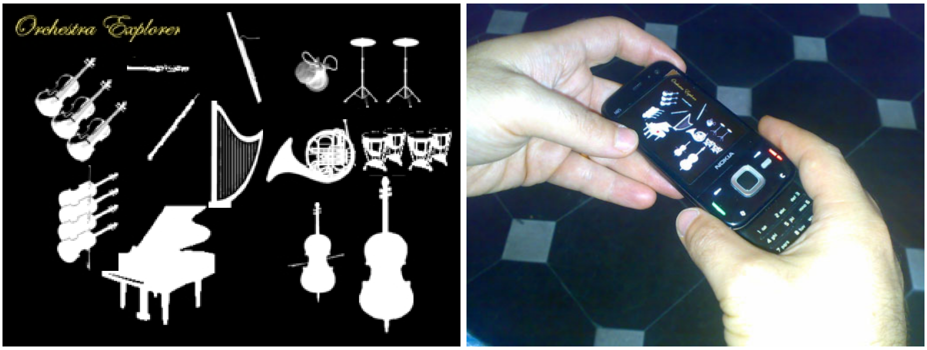

Fig. 2. The *Mobile Orchestra Explorer*. On the left the orchestra space that the user can see on the display of the mobile phone. On the right, an user trying the application.

Sync'n'Move[3] [7] enables users to experience a novel form of social interaction based on music and gesture, using mobile phones and the SAME platform. Users move rhythmically (e.g., dance) wearing their mobiles. Their phase synchronization, extracted from their gestures, is measured and used to modify in real-time the performance of a pre-recorded music. This is a first example of shared collaborative active music listening experience. Every time the users are successful in the synchronization task, the music orchestration and rendering is enhanced; whereas in cases of low synchronization, i.e., poor collaborative interaction, the music gradually corrupts, looses sections and rendering features, until it becomes a very poor monophonic audio signal.

Two sample applications are devoted to real-time control by mobile devices of expressivity in music. In the first one, a mobile phone is used for controlling the emotional expression of ringtones. The user chooses an emotion for his/her ringtone. The ringtone is sent to a server where it is processed using the KTH performance system for expressive music performance [8] and returned to the user's handset with the desired emotional expression. The KTH performance system controls different aspects of the performance, such as tempo, dynamics, articulation, orchestration, by associating pre-assigned values for each emotion[4]. In the second one, *pyDM* is used for expressive control of a piano performance[5]. A computer-controlled piano is connected to a computer running pyDM. This is a program for interactive control of expressivity in music performance, using the KTH rule system for music performance (see Figure 3). Again, each rule controls different aspects of the performance, such as tempo, dynamics, and articulation. Rule values can be adjusted separately, or mapped

[3] Contributors: Giovanna Varni, Paolo Coletta, Gualtiero Volpe, Antonio Camurri, Corrado Canepa, Maurizio Mancini, Barbara Mazzarino – DIST.

[4] Contributors: Roberto Bresin – KTH; Jarno Seppanen – Nokia.

[5] Contributors: Marco Fabiani, Roberto Bresin, Gaël Dubus – KTH.

to more intuitive control parameters, such as the Activity-Valence space, in which different basic emotions can be expressed (e.g., happiness, sadness, tenderness, anger). In pyDM, the Activity-Valence value is shown by a moving circle, whose color and dimension vary according to the expressed emotion (see Figure 3). The program can be controlled using a mobile phone graphical interface, or by tilting the phone, as well as by shaking it in different ways to express different emotions.

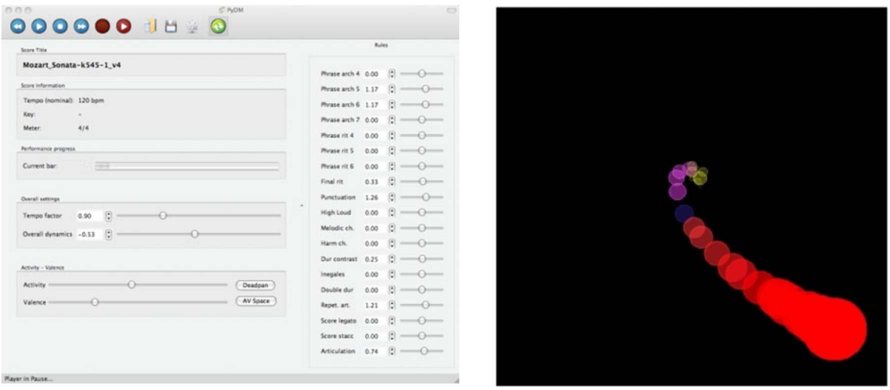

Fig. 3. On the left *PyDM*. On the right, representation of trajectories in the Activity-Valence space as moving circles, whose color and dimension vary according to the expressed emotion.

Context-awareness is particularly addressed in *Zagora*[6], a context-aware mobile music player, which detects the ambient situation using audio analysis and retrieves a playlist of suitable music. The Zagora player is doing advanced audio processing to differentiate between situations like *street, restaurant, car, office,* and *meeting,* and uses the situation information to filter down an online music catalog. The user can see the current audio analysis results, generate a playlist online, and start streaming music. Finally, all resulting playlists can be browsed for other similar online music.

Further sample applications adopt a game-like paradigm. The *Mobile Sonic Playground*[7] demonstrates examples of individual and collective sonic applications using mobile phones as musical instruments and sounding toys. The user interacts with the phone accelerometers and keypad keys, and generates control events that are captured, processed, and rendered to sound using the phone embedded Mobilophone framework. Several playing styles and synthesized sound selections are available.

The *Fishing Game*[8] illustrates novel technologies on gestural sound control and embodied active listening. The system makes use of gesture recognition and analysis, driving a sound engine. When users with their mobile phones mimic gestures such as pouring a glass or brushing teeth, they can listen to the sound associated to such gestures. This illustrates emerging uses of embedded sensors in mobile phones.

---

[6] Contributors: Antti Eronen, Jussi Leppänen, Jarno Seppänen – Nokia.

[7] Contributors: Jari Kleimola, Sami Oksanen, Vesa Välimäki – TKK.

[8] Contributors: Pierre Jodlowski, Baptiste Caramiaux, Grace Leslie, Norbert Schnell, Diemo Schwarz, Bruno Zamborlin, Frédéric Bevilacqua, Hugues Vinet, Olivier Warusfel – IRCAM.

Finally, the *Grain Stick* installation[9] offers a collaborative interactive experience featuring music by Pierre Jodlowski. One or two participants shake a virtual tube by means of two manual sensors that set off a waterfall of sound grains (like a rain stick) in a sound space spatialized with WFS technology. The sounds of the grains, generated by the corpus-based synthesis engine CataRT, overlap the surrounding soundscape and percussive sounds that are triggered by the users' movements. The virtual stick can be used by one person alone with both hands or by two users, thus including a social dimension (see Figure 4). Beyond technical feasibility, an important aspect of this work on the artistic side has been to experience a new kind of interactive musical form.

Fig. 4. Two users experiencing the *Grain Stick* installation

3 Evaluation

Evaluation has been carried out by asking participants to fill questionnaires. These included general questions, concerning the overall evaluation of the active listening concept, and questions for the sample applications. Moreover, the evaluation of some of the sample applications included specific questions especially devoted to them.

The general questions concerning all the prototypes are reported in Figure 5. The questions that participants answered for each sample application are reported in Figure 6. The questionnaire also included information about age, gender, musical skills and habits (e.g., preferred music genre, time spent in listening to music, etc.).

Evaluation involved 108 participants. 82 attended public sessions; 16 attended a special session dedicated to expert users (music professionals). Table 1 shows the results for the questions in Figure 6. Answers were collected using 11 point scale

9 Contributors: Pierre Jodlowski, Grace Leslie, Markus Noisternig, Norbert Schnell, Joseph Sanson, Diemo Schwarz, Bruno Zamborlin, Frédéric Bevilacqua, Hugues Vinet, Olivier Warusfel – IRCAM.

from -5 to 5. The table include the number of subjects that answered the question for a specific application (N), and mean and standard deviation of the results.

Q1: What did you expect from this experience (check all that apply) ?
□Have fun
□Learn
□New music experience
□ Better communicate with peers
Other ______________________________

Q2: What was you first impression ?
Very negative □□□□□□□□□□□ Very positive

Q2: The strength of your experience was:
Very weak □□□□□□□□□□□ Very strong

Q3: Which of the following areas do you think could benefit from the project (check all that apply) ?
□New entertainment
□New technology
□New form of art
□ Kinesthetic/motor abilities
□Ability to communicate
□For no good use at all
Other ______________________________

Fig. 5. Evaluation questionnaire: the general questions on the active listening experience

Q1: How easy is it to understand how the application works?
Very difficult □□□□□□□□□□□ Very easy

Q2: How much do you feel in control of the application?
Very little □□□□□□□□□□□ Very much

Q3: How do you find the level of interaction?
Low □□□□□□□□□□□ High

Q4: What do you think about this application?
Boring □□□□□□□□□□□ Funny
Uninteresting □□□□□□□□□□□ Ineresting
Nothing for the future□□□□□□□□□□□ Something for the future
Not engaging □□□□□□□□□□□ Engaging
I did not enjoy it □□□□□□□□□□□ I enjoyed it

Fig. 6. Evaluation questionnaire: questions for each sample application

Table 1. Results from user evaluation for the SAME sample application presented at the Agora Festival, Paris, France, June 2009. Questions Q1, Q2, Q3, and Q4 are reported in Figure 6. Answers were collected using a 11 point scale from -5 to 5. The number of participants (N), average values ($\bar{x}$) and standard deviations (σ) are reported for the both session types: i.e., with expert and non-expert users.

Question		Audio Explorer						Fishing Game						Grain Stick						Mobile Orchestra Explorer						Sync'n'Move					
		Experts			Non-experts			Experts			Non-experts			Experts			Non-experts			Experts			Non-experts			Experts			Non-experts		
		N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ
Q1	Very difficult vs. Very easy	11	2.45	1.51	18	2.22	2.62	10	4.5	0.71	15	3.53	2.33	12	3.33	1.15	24	1.54	2.21	10	3.9	1.52	13	4.31	1.18	9	3.56	1.74	10	4.3	1.57
Q2	Very little vs. Very much	9	2.89	1.05	15	1.87	1.81	10	3.3	1.7	13	2.46	1.66	13	3.23	2.09	20	1.95	2.31	8	1	1.93	13	4.15	0.8	8	0.38	2.62	9	2.11	1.83
Q3	Low vs. High	11	1.91	1.87	16	2.5	1.71	10	3.5	1.18	14	2.86	2.44	13	3.54	1.61	22	3.86	1.67	9	0.11	3.06	13	2.54	1.98	9	0.78	3.03	10	1.7	1.77
Q4	a) Boring vs. Funny	11	1.18	2.14	16	2.38	1.82	10	3.9	1.1	15	3.67	1.91	13	3.69	1.18	24	3.29	1.73	9	0.11	2.62	14	2.14	2.28	10	0.7	3.02	10	2	3.27
	b) Uninteresting vs. Interesting	11	1.73	2.41	17	3.06	1.75	10	2.7	1.95	14	2.43	1.79	13	3.62	1.19	24	3.96	1.65	9	0.78	3.19	13	1.77	1.24	10	-1	3.13	10	0.7	2.54
	c) Nothing for the future vs. Something for the future	11	1.27	3.17	16	2.63	2.16	10	2	3.16	14	2	2.66	13	3.69	1.49	24	3.79	1.86	9	1.33	3.39	14	2.86	1.75	10	-0.2	3.19	10	1.5	2.76
	d) Not engaging vs. Very engaging	11	1.36	2.54	16	1.56	2.13	10	1.8	1.99	13	2.46	2.15	13	3.85	1.07	24	3.42	1.84	9	-0.33	3.12	13	1.92	1.89	10	0.1	3.03	10	0.9	2.33
	e) did not enjoy it at all vs. I enjoyed it very much	11	0.82	2.04	16	2.38	1.71	10	2.1	1.79	14	3.07	1.94	13	3.77	1.24	24	3.54	2.11	9	0.11	2.98	13	2.38	1.5	10	-0.3	2.63	10	1.4	2.67

Question		pyDM						Emotional expression of ringtones						Zagora						Mobile Sonic Playground					
		Experts			Non-experts			Experts			Non-experts			Experts			Non-experts			Experts			Non-experts		
		N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ	N	$\bar{x}$	σ
Q1	Very difficult vs. Very easy	12	4.5	0.8	23	3.61	1.75	10	3.9	1.29	15	3.2	1.93	13	4.69	0.63	11	2.55	2.54	12	3.75	1.42	10	2.8	1.87
Q2	Very little vs. Very much	11	1.73	3.07	21	3.05	2.01	10	2.1	2.6	13	2.54	1.71	10	-0.7	2.41	10	1.5	2.51	11	0.91	2.74	8	2.63	1.41
Q3	Low vs. High	12	0.92	3.48	23	2.91	1.68	10	-0.3	3.23	14	1.36	1.98	13	-0.85	2.54	11	0.91	2.12	12	-0.92	2.97	9	2.67	1.73
Q4	a) Boring vs. Funny	12	1.75	3.05	23	3.26	1.79	10	-1.2	2.82	15	1.8	2.37	13	-0.92	2.5	11	1.73	3.17	12	-0.83	2.48	10	2.3	3.23
	b) Uninteresting vs. Interesting	12	0.42	3.2	23	3.13	1.66	10	-1.3	2.67	15	1.53	2.07	13	-0.92	2.66	10	2	2.79	11	-0.27	2.1	10	2	3.02
	c) Nothing for the future vs. Something for the future	12	1.67	2.67	23	2.35	2.04	10	-1.7	3.3	15	1.8	1.93	12	0.25	3.19	10	2.6	1.43	11	-0.27	2.69	10	2.6	2.37
	d) Not engaging vs. Very engaging	12	0.75	3.14	22	2.36	1.5	10	-2	2.45	14	1.36	2.53	12	-0.58	2.78	9	1.22	2.54	12	-1.08	2.5	10	2.4	2.46
	e) did not enjoy it at all vs. I enjoyed it very much	12	0.92	3.03	22	2.95	1.59	10	-1.5	3.27	14	1.79	1.85	12	-0.08	2.71	10	2	1.49	12	-0.58	2.27	10	2.3	2.5

Based on such feedback, we can infer that visitors generally liked the SAME applications. Feedback from both groups included valuable criticism, suggestions, and proposals for improvements. The demonstrated applications have a potential for the future, but more research and development work needs to be done. Indeed, some of the applications were finished and well refined installations (e.g., Grain Stick), whereas others were rather proof-of-concepts that still need to be further developed (e.g., Synch'n'Move). Such different development stages of the applications may have affected the average appreciation by participants.

Evaluation pointed out that the users are anxious to see social networking features implemented as a part of the applications. This is encouraging for future research on embodied social interaction envisaged in SAME. In general, the interaction between users and personal file sharing was appreciated. Merging of the applications or combining them into more versatile systems was also suggested.

Finally, if from the one hand, the applications received an overall positive feedback by both expert and non-expert users, on the other hand, non-expert users especially appreciated active listening both as a new way for listening to music and also as an educational tool for gaining a better understanding of how music is made, structured, and performed.

4 Conclusions

This paper presented a survey of the sample applications the SAME EU Project presented at Agora Festival (IRCAM, Paris, June 2009), as mid-term milestone of its research and development work. We believe that such applications represent a useful test-bed for future paradigms of active experience and User-Centric Media.

The applications have been evaluated by expert and non-expert participants, whose feedback will be used for refining the requirements of the project and for moving towards the final set of comprehensive prototypes of systems and applications for context-aware mobile social and active listening to music.

Acknowledgments

This work has been partially supported by the EU-ICT Project SAME. We thank all the contributors to the sample applications for their precious work.

References

1. Camurri, A., Canepa, C., Volpe, G.: Active listening to a virtual orchestra through an expressive gestural interface: The Orchestra Explorer. In: Proceedings 2007 Intl. Conference on New Interfaces for Musical Expression (NIME 2007), New York, USA (June 2007)
2. 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 2008 Intl. Conference on New Interfaces for Musical Expression (NIME 2008), Genova (2008)

3. Karjalainen, M., Maki-Patola, T., Kanerva, A., Huovilainen, A.: Virtual air guitar. Journal of the Audio Engineering Society 54(10), 964–980 (2006)
4. Pakarinen, J., Puputti, T., Valimaki, V.: Virtual slide guitar. Computer Music Journal 32(3), 42–54 (Fall 2008)
5. Camurri, A., Coletta, P., Demurtas, M., Peri, M., Ricci, A., Sagoleo, R., Simonetti, M., Varni, G., Volpe, G.: A Platform for Real-Time Multimodal Processing. In: Proceedings International Conference Sound and Music Computing 2007 (SMC 2007), Lefkada, Greece (2007)
6. Vinyes, M., Bonada, J., Loscos, A.: Demixing Commercial Music Productions via Human-Assisted Time-Frequency Masking. In: 120th AES Convention, Paris (2006)
7. Varni, G., Mancini, M., Volpe, G.: Sync'n'Move: social interaction based on music and gesture. In: Proc. 1st International ICST Conference on User Centric Media, Venice (2009)
8. Friberg, A.: pDM: an expressive sequencer with real-time control of the KTH music performance rules movements. Computer Music Journal 30(1), 37–48 (2006)