User-Centered Control of Audio and Visual Expressive Feedback by Full-Body Movements

Ginevra Castellano[1], Roberto Bresin[2], Antonio Camurri[1], Gualtiero Volpe[1]

[1] InfoMus Lab, DIST - University of Genova
Viale Causa 13, I-16145, Genova, Italy
{Ginevra.Castellano, Antonio.Camurri, Gualtiero.Volpe}@unige.it
[2] KTH, CSC School of Computer Science and Communication,
Dept. of Speech Music and Hearing, Stockholm
roberto@kth.se

Abstract. In this paper we describe a system allowing users to express themselves through their full-body movement and gesture and to control in real-time the generation of an audio-visual feedback. The systems analyses in real-time the user's full-body movement and gesture, extracts expressive motion features and maps the values of the expressive motion features onto real-time control of acoustic parameters for rendering a music performance. At the same time, a visual feedback generated in real-time is projected on a screen in front of the users with their coloured silhouette, depending on the emotion their movement communicates. Human movement analysis and visual feedback generation were done with the EyesWeb software platform and the music performance rendering with pDM. Evaluation tests were done with human participants to test the usability of the interface and the effectiveness of the design.

Keywords: affective interaction; expressive gesture; multimodal environments; interactive music systems

1 Introduction

In human-computer interaction there is an increasing attention on the identification of new paradigms of expressive human-machine interaction, and on the design of interfaces combining information from several channels in a multimodal way. New results in these fields would lead to fully use the power and impact of human non-verbal expressive communication in experience-centric tasks and collaborative applications. Endowing computers with the ability of analysing, interpreting, and reacting to users' expressive gestures and behaviour is still an open issue ([1],[2],[3]).

Several studies in psychology of emotion show that the body is an important channel to communicate affect and that people naturally express emotions through the use of their motor skills (see for example [4],[5],[6],[7],[8],[9]). The focus of our research is on the non-verbal aspects of the communication process between humans and computers and in particular on forms of expressing emotions based on the use of human movements and gestures [10].

In this paper we focus on the role of expressive intentions in the field of interactive music systems, i.e. systems able to process expressive gesture for the generation and control of matching musical signals ([11],[12]). The work presented here could also seen as an artistic application, since it also involves real-time generation of graphics and music performance driven by computer vision techniques [13]. However, the main focus of the paper is on how the user interacts with the environment with her body when multimodal feedback is provided, taking into account theories of embodied music cognition [14]. In this context, we want to develop interaction strategies which involve full-body movements and gestures at different levels of abstraction in an environment that can evolve and dialogue with the user. This leads to Multimodal Environments (MEs) which allow for creative, multimodal user interaction, by exhibiting real-time adaptive behaviour. We refer in particular to immersive environments enabling communication by means of full body movement, such as dance and expressive gestures, singing, and music playing. Users get a feedback from the environment in real-time in terms of sound, music, visual media, and changes in the environment.

The work presented in this paper is a contribution to the investigation of MEs with a special focus on expressive and emotional communication. We propose a system in which users can express themselves through their full-body movements and gestures which are used for controlling the real-time generation of an audio-visual feedback. The system analyses in real-time the user's full-body movements and gestures, extracts expressive motion features and classifies movement according to five different emotions (sadness, serenity, happiness, fear and anger). The values of the expressive motion features are mapped onto real-time control of acoustic parameters for rendering the expressivity in a musical performance. At the same time, the output of the classification drives the real-time generation of a visual feedback: users can see their silhouette projected on a screen in front of them, with different colours depending on the emotion communicated by their movements. This leads to the generation of an affective interaction, where users are immersed in an affective loop with the system [15].

Evaluation tests with subjects were done to evaluate the effectiveness of the designed interaction. During the evaluation phase, we addressed the following questions: (1) Is the designed audio-visual feedback suitable to reproduce the expressivity of the movement performed by subjects in the space? (2) Do the subjects understand the interaction? (3) Do subject feel in control of the musical and visual feedback? Do they become 'performers'? (4) Are there differences, at the perceptive level, between the visual and audio feedback?

This paper is divided into two main parts. The first part focuses on the design of the application. The second part deals with the usability evaluation with human participants, trying to give some answers to the above questions.

2 Overview of the system

The system presented in this paper is based on the integration of two different systems: EyesWeb [16], developed at DIST InfoMus Lab, and pDM (see [17],[18]), developed at KTH. EyesWeb is an open software platform for multimodal analysis and the development of interactive systems and MEs. pDM is a program for the real-time rendering of expressive music performance, allowing for different emotional characterizations by manipulating acoustic parameters such as sound level, tempo, and articulation.

The system integrating EyesWeb and pDM acquires input from a video camera, processes information related to the expressivity of human full-body movements, extracts expressive indicators, classifies movements in terms of emotions (sadness, serenity, happiness, fear and anger) and controls the expressive rendering of music performance and the visual feedback generation (see Figure 1).

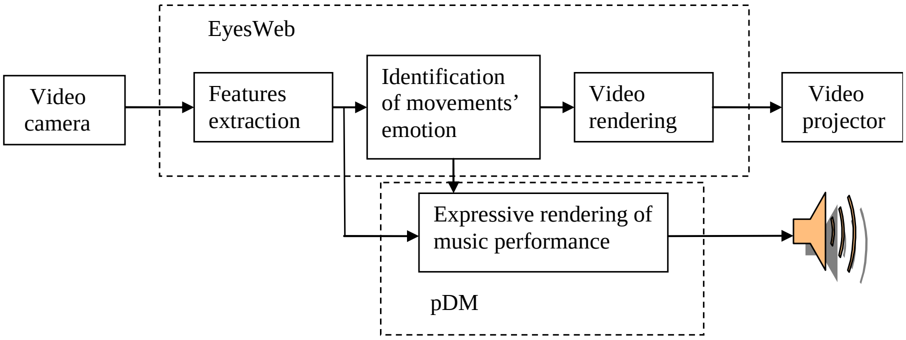
EyesWeb
Video camera
Features extraction
Identification of movements' emotion
Video rendering
Video projector
Expressive rendering of music performance
pDM

Fig.1. Overview of the system.

Human movement is being analysed with the EyesWeb platform and expressive motion features are being automatically extracted with the support of the EyesWeb Expressive Gesture Processing Library [19].Visual feedback generation was done with EyesWeb, and the music performance rendering with pDM.

3 Analysis of human full-body movement

Movement and gesture expressivity is a key element both in understanding and responding to users' behaviour. Several studies showed that the body is an important channel to communicate expressivity and people naturally express emotions through the use of their motor skills ([4],[5],[6],[7],[8],[9],[20]).

We focused on two expressive motion features: the Quantity of Motion (QoM) and the Contraction Index (CI). Both cues are global indicators of human movements from energy (QoM) and space (CI) perspectives.

QoM is an approximation of the amount of detected movement, based on Silhouette Motion Images (SMIs). A SMI is an image carrying information about variations of the silhouette shape and position in the last few frames.

$$SMI[t] = \left\{ \sum_{i=0}^{n} Silhouette[t-i] \right\} - Silhouette[t].$$

The SMI at frame t is generated by adding together the silhouettes extracted in the previous n frames and then subtracting the silhouette at frame t. The resulting image contains just the variations happened in the previous frames. QoM is computed as the area (i.e., number of pixels) of a SMI divided by the area of the current silhouette, i.e. normalized in order to obtain a value usually ranging from 0 to 1. It can be considered as an overall measure of the amount of detected motion, involving velocity and force.

$$QoM = \frac{Area(SMI[t,n])}{Area(Silhouette[t])}.$$

CI is a measure, ranging from 0 to 1, of how the user's body expands or contracts in the surrounding space. It can be calculated using a technique related to the bounding region, i.e., the minimum rectangle surrounding the user's body: the algorithm compares the area covered by this rectangle with the area currently covered by the silhouette. These two features are extracted and processed in order to control music and visual outputs. We are exploring relations between the emotions and other movement characteristics such as fluidity, impulsiveness, direction changes, amount of upward movements, symmetry and repetitivity.

4 Visual feedback generation based on emotional communication

Visual feedback was designed to respond to the user's expressive motor behaviour. The user moves in a large space in front of the camera. The system identifies user's gestures and projects her silhouette on a wall in front of her by using different colours, which depend on the identified movement's expressivity. We focused on five emotions, each of them related to one specific combination of QoM and CI values: sadness, serenity, happiness, fear and anger.

According to theories from psychology on emotion in movements ([4],[5],[6],[7],[8],[9]) and from previous studies on expressivity ([10],[20],[21],[22]), we defined the following correspondence between emotion and movement characteristics (Table 1):

Table 1. Relations between emotion and movement characteristics

Emotion	Movement characteristics
Sadness	Slow (low QoM), contracted (high CI)
Serenity	Slow (low QoM), expanded (low CI)
Happiness	Fast (high QoM), expanded (low CI)
Anger	Very fast (very high QoM), expanded (low CI)
Fear	Fast (high QoM), contracted (high CI)

Applying results from previous studies on the association of colours to expressive music performance [23], we associated specific colours to the emotions communicated by the user's movement: violet to sadness, pink to serenity, yellow to happiness, red to anger, and blue to fear.

Figure 2 shows examples of the visual feedback as it was presented to the user.

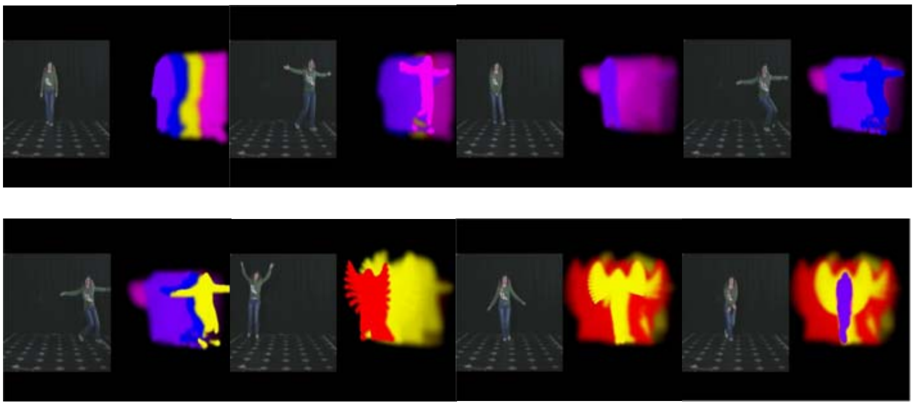

Fig.2. The coloured silhouettes are examples of visual feedback as presented to the user. The current instant of time corresponds to the silhouette. Previous movements with different expressivity are still visible as traces left by the silhouette. The current input frame is shown on the left-hand side of each silhouette.

5 Acoustic feedback generation

Acoustic feedback was designed using pDM. It is a consolidated knowledge that sound level, tempo, and articulation accounts for more than 90% of the emotional expression in music performance (for an overview see [24]).

The values of QoM and CI provided by EyesWeb are used for controlling the real-time expressive performance of a musical score. The user can freely choose which score to control with her body gestures. QoM and CI are mapped onto tempo and sound level respectively. In this way the user's performance gets faster for faster movements (i.e. higher QoM values) and louder when the user is expanding her body (i.e. lower CI). Articulation is controlled by mapping the emotion detected in the user's movements to the articulation parameters in pDM. More *legato* articulation for emotions with lower activity (low QoM and contracted movements) and more *staccato* articulation for those characterized by higher energy (high QoM and expanded movements). Furthermore, articulation related to angry body movements was played less *staccato* than in happy movements.

More in detail, the overall expressivity in pDM was set to a set-up corresponding to a musically neutral performance (i.e. with values of phrasing, articulation, tempo and sound level corresponding to those typical of a neutral performance). It is the user with her QoM and CI who directly controls tempo and sound level. These two acoustic parameters have been demonstrated to be the most important, together with articulation, for the communication of emotion in expressive music performance [25]. Therefore the emotion in the resulting performance is a direct mapping of user's full body emotional behaviour in the space in front of the camera.

6 Usability evaluation tests with subjects

An evaluation test was conducted in a lab setting with subjects for testing the usability of the interface and the effectiveness of the design. The questions which we addressed in this phase were the following: (1) Is the designed audio-visual feedback suitable to reproduce the expressivity of the movement performed by subjects in the space? (2) Do the subjects understand the interaction? (3) Do subject feel in control of the sound/colour generation? Do they become 'performers'? (4) Are there differences, at the perceptive level, between the visual and the audio feedback?

6.1 Subjects

A group of fourteen people (six male and eight female, from 20 to 51 years old, average age: 26.78), participated to the experiment. Each of them tried the system for the first time during the experiment.

6.2 Set up

The experiment was realized in a square room with diffuse lights, in order to allow the system to track the participants' body. A computer running EyesWeb and pDM was connected through an Edirol FA-101 sound card to wireless headphones, which were worn by the subjects. A video camera Panasonic WVCP450/G, located in front of the users, was used for capturing the full-body movement of the participants and for providing EyesWeb with the visual input. 25fps non interlaced were used. A video-projector projected the visual feedback on the wall in front of the subjects. A second video camera Sony MiniDV "DCR PC1000E" / 3CMOS was used for recording possible comments of the participants as well as their movements from a frontal-lateral view.

6.3 Procedure

Subjects were asked to move in the space at disposal (a square with dimensions 2.80m X 3.60m, 1.75m far from the frontal video camera) and they were informed that they would be able to change audio and visual feedback by moving. They could hear the audio feedback with headphones and see the visual feedback on a wall in front of them. Subjects had the possibility to choose between two pieces of music to be rendered through their movements: 'We are the champions' by Queen and the Trio Sonata n. 3 by Johann Sebastian Bach. Six of them chose the Trio Sonata by Bach and eight the Queen's song. Subjects did not listen to the chosen piece before starting the experiment.

Participants were asked to move and to try obtaining a visual and audio feedback which they liked. That choice was done in order to induce them to be active and to allow them to achieve a full experience of the application. They were told that they were free to test the application as long as they wanted and they might signal the end of their trial by leaving the space in front of the video camera.

6.4 Methods for evaluation

A triangulation between methods was used to evaluate the application; notes of the experimenter, video recordings, and interviews to the subjects were compared.
During the test, participants' behaviour was annotated by the experiments. Such notes were used after the test to compare results from the off-line analysis of the video recordings. Participants were interviewed at the end of the experiment. They had to answer the following questions:

1. Overall, what did you think about the system? What did you do? How?
2. Did the audio change by your movement? How?
3. Did the visual effects change by your movement? How?
4. Did you feel in control of the music interpretation?
5. Did you feel in control of the colours generation?
6. Do you think expressivity of your movement played an important role in the generation of the audio and visual feedback?
7. Did you feel immersed in the environment?

6.5 Results

From the qualitative analysis of the recordings and the notes, we can highlight some general observations.

1. People need a certain amount of time to become familiar with the application: the average time they needed was near to the four minutes.
2. Only the most active people obtained a full experience of the application.
3. Only one person talked during the test: sentences like “I want to become yellow!” and “Oh, perfect, done!” demonstrate the attempt to interact with the system to achieve some specific goal.
4. The most part of people (71.4%) continued looking at the wall with the visual effects; people that focused on the audio, did not make use of the arms while interacting with the interface.

We report in the following results from the analysis of the answers given by the subjects during the interviews.

1. Overall, what did you think about the system? What did you do? How?

Participants' main goal was to understand how the system worked. One person (student in philosophy) observed that the application was interesting from the artistic perception point of view and it was a good example of non-passive art: ‘the presence of rules in the application makes the subject open to the world’.

2. Did the audio change by your movement? How?

Eleven people (78.6%) thought their movement changed the audio. Between them, one person understood the whole mapping, four understood the relationship between quantity of motion/velocity and tempo in music, six didn’t understand how the mapping worked.

3. Did the visual effects change by your movement? How?

Twelve people (85.7%) thought their movement changed the visual effects. Between them, one person understood the whole mapping (the same who understood the movement-audio mapping), six understood the relationship between quantity of motion/velocity and colours generation and five didn’t understand how the mapping worked. Only one person thought the movement didn’t change both audio and video.

4. *Did you feel in control of the music interpretation?*

The 7% of the participants felt in control of the music interpretation, the 36% felt in control in an acceptable way, the 57% didn’t feel in control.

5. Did you feel in control of the colours generation?

The 28.6% of the participants felt in control of the colours generation, the 42.8% felt in control in an acceptable way, the 28.6% didn’t feel in control.

6. Do you think expressivity of your movement played an important role in the generation of the audio and visual feedback?

The 35.7% of the participants answered *Yes*, the 35.7% answered *Enough*, the 28.6% answered *No*.

7. Did you feel immersed in the environment?

The 71.4% of the participants answered *Yes*, the 14.3% answered *Enough*, the 14.3% answered *No*.

7 Discussion

From the analysis of users’ behaviour and their direct feedback we conclude that our system provided strong sensations of participation, interaction and immersion.

Most part of users perceived the changes in audio and colours generated by their movement and many of them understood the mappings we designed, with a prevalence for the understanding of the role played by the quantity of motion in colours generation and music performance rendering rather then by the contraction index. Nobody found out relationships between quantity of motion and articulation of the music performance. Further, most part of the users felt in control of the colours generation, while the control of the music rendering is less perceived. From the perceptive point of view, the tests highlighted the major role played by the visual channel. Users seem to understand better the effects which their movements have on the colour generation. The analysis of their answers highlights that they feel more in control of the visual feedback generation than of the audio feedback. This can be due to the fact that maybe some of the users were not used to listen to or did not know the music used during the experiment. Future tests will be designed in such a way that the participants get to know the music before starting the experiment, so that they can fully experience the system.

As it emerged from the interviews, users seem to understand the role that the expressivity plays in the interaction. Off-line analysis of the recorded videos demonstrates that the system provides users with a new way of expression, allowing them to communicate and interact with a computer in real-time and to obtain an appropriate feedback.

8 Conclusion

In this paper we presented a system allowing users to use their own body as an expressive interface, allowing the control of music expressivity with the body. Body expressivity is also used as channel for the communication of emotional, affective content: visual media generation and control become part of an affective analysis and synthesis loop involving the user in a perception-action process.

At the following link, ftp://infomus.dist.unige.it/Pub/ACII2007/, it is possible to see a concrete example of how the users can interact with the system.

Currently, our system is used for entertainment task and has been tested in a national exhibition. More generally, the results from this experiment confirm the general acceptance and engagement of users for experiences of participative listening [10] and visual affective interaction. This seems a promising perspective for future scenarios where the communication human-computer is based on non-verbal emotional channels. In order to further develop and validate the proposed system, in future experiments we will test it by giving mismatching feedback to the user for example by providing mismatching visual and/or audio feedback to the emotion of the user's body motion.

Acknowledgements. The research work has been realised in the framework of the EU-IST Project HUMAINE (Human-Machine Interaction Network on Emotion), a Network of Excellence (NoE) in the EU 6th Framework Programme (2004-2007) and the COST287 Action ConGAS (Gesture CONtrolled Audio Systems).

References

1. Picard, R.: Affective Computing. Boston, MA: MIT Press, 1997.

2. Hashimoto, S.: KANSEI as the Third Target of Information Processing and Related Topics in Japan. In Proc. International Workshop on KANSEI: The technology of emotion, Genova, 1997, pp. 101-104.
3. Cowie, R., Douglas-Cowie E., Tsapatsoulis N., Votsis N., Kollias N., Fellenz W., Taylor J.: Emotion Recognition in Human-Computer Interaction. IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 32-80, 2001.
4. Scherer, K.R., and Wallbott, H.G.: Analysis of Nonverbal Behavior. HANDBOOK OF DISCOURSE: ANALYSIS, Vol. 2, Cap.11, Academic Press London, 1985.
5. Wallbott, H.G., and Scherer, K.R.: Cues and Channels in Emotion Recognition. Journal of Personality and Social Psychology, 1986, Vol.51, No.4, 690-699.
6. DeMeijer, M.: The contribution of general features of body movement to the attribution of emotions. Journal of Nonverbal Behavior (13), 247 – 268, 1989.
7. Wallbott, H.G.: Bodily expression of emotion. European Journal of Social Psychology, Eur. J. Soc. Psychol. 28, 879-896, 1998.
8. 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.
9. 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, pp. 20-39, Springer Verlag, 2004.
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, 12(1):43- 53, IEEE CS Press, 2005.
11. Camurri, A.: Interactive Dance/Music Systems, Proc. Intl. Computer Music Conference ICMC-95, pp.245-252, The Banff Centre for the arts, Sept.3-7, 1995, Canada, ICMA-Intl.Comp.Mus.Association.
12. Camurri, A. and Trocca, R.: Movement and gesture in intelligent interactive music systems. In M.Battier and M.Wanderley (Eds.), Trends in Gestural Control of Music, Ircam Publ., 2000.
13. Krueger, M: Artificial Reality II. Addison-Wesley Professional, 1991
14. Leman, M: Embodied Music Cognition and Mediation Technology, Cambridge, MA: The MIT-Press (in print).
15. Höök, K: User-Centred Design and Evaluation of Affective Interfaces, In From Brows to Trust: Evaluating Embodied Conversational Agents, Edited by Zsofia Ruttkay and Catherine Pelachaud, Published in Kluwer's Human-Computer Interaction Series, 2004.
16. Camurri A., Coletta P., Massari A., Mazzarino B., Peri M., Ricchetti M., Ricci A., Volpe G.: Toward real-time multimodal processing: EyesWeb 4.0, in Proc. AISB 2004 Convention: Motion, Emotion and Cognition, Leeds, UK, March 2004.
17. Friberg, A.: pDM: an expressive sequencer with real-time control of the KTH music performance rules. Computer Music Journal, 30(1), 37-48, 2006.
18. Friberg, A., Bresin, R. and Sundberg, J.: Overview of the KTH rule system for music performance. Advances in Experimental Psychology, special issue on Music Performance, 2(2-3), 145-161, 2006.
19. Camurri, A., Mazzarino, B., and 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, 2004.
20. Camurri, A., Lagerlöf, I, Volpe, G.: Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques, International Journal of Human-Computer Studies, 59(1-2), pp. 213-225, Elsevier Science, July 2003.
21. Camurri, A., Castellano, G., Ricchetti, M., and Volpe, G.: Subject interfaces: measuring bodily activation during an emotional experience of music. In S. Gibet, N. Courty, J.F.

Kamp (Eds.), Gesture in Human-Computer Interaction and Simulation, Volume 3881, pp. 268-279, Springer Verlag, 2006.

22. Dahl, S. and Friberg, A.: (forthcoming). Visual perception of expressiveness in musicians' body movements. Music Perception.
23. Bresin, R.: ‘What is the color of that music performance?’, In proceedings of the International Computer Music Conference - ICMC 2005, 367–370, 2005.
24. Juslin, P. and laukka, P.: Communication of emotions in vocal expression and music performance: Different channels, same code?, Psychological Bulletin, 129:5, pp. 770-814, 2003.
25. Juslin, P.N.: Communicating Emotion in Music Performance: a Review and Theoretical Framework. In: Juslin, P.N. & Sloboda, J.A. (eds.): Music and Emotion: Theory and research. Oxford: University Press. (2001) 309-337.