An Emotion-Driven Interactive System

Francisco Ventura, António Pedro Oliveira and Amílcar Cardoso

Centre for Informatics and Systems, University of Coimbra, Portugal
{fventura, apsimoes}@student.dei.uc.pt, amilcar@dei.uc.pt

Abstract. Emotion-based interactive music systems have great application potential, namely in entertainment and healthcare. This paper describes an installation that explores the interactive capabilities of EDME, an emotion-based music engine we are developing. At the core of the installation there is an affective computer system that selects appropriate music and images to express its emotional state. The installation allows people to experience and influence the emotional behavior of this system. We made two experiments where people were able to ascribe emotions to the system in a natural way, relating arousal with the amount of movement and valence with the number of presences, as expected. The system transmitted expected emotions and gains from using both media (music and images) to express its emotional state.

Keywords: Interactive music system, automatic music production, emotions and music.

1 Introduction

Building interactive systems with the capability to produce music with an intentional emotional content involves multidisciplinary competences in Affective Computing [20], Human Computer Interaction, Music Psychology and a mixture of these domains [19]. The interaction may involve the perception of emotions from users [3, 8, 26]. This can be done with sensors to recognize psychophysiological signals, facial expressions, gestures and others [1, 7, 17]. The system must be able to adapt its emotional state to the perceived information and express it in some human recognizable way. The expression of its emotional state may be performed by suitably adapting media like music and images.

We intend to explore the capability of an affective system in the expression of a desired emotion to an environment [3, 8, 18, 25, 26], in such a way that people may recognize system emotional state in a natural way. We are developing a solution with a high degree of adaptability due to the flexibility of 2 auxiliary structures (music base and knowledge base) being used in EDME, an emotion-based music engine that intends to control the emotional expression of produced music, that makes the system a good choice for almost every context. This paper starts with a review of systems related to our work (section 2). Then, we describe our emotion-based interactive installation (section 4) that intends to interact with people by using music and images. This installation intends to explore the interactive capabilities of EDME (section 3).

The installation allows people to experience and influence the emotional behavior of the computer system. The description of the two experiments made to test this system is presented in section 5. Section 6 analysis results from the second experiment and, finally, section 7 draws some conclusions.

2 Related work

The emotional interaction with people by using different media is a promising field of research. The importance of developing systems with such a capability is evident to the society. They can be used in contexts where there is a need to create environments capable of inducing certain emotional experiences: intelligent spaces [26], virtual environments [3, 15, 21], healthcare spaces [16], dance spaces [25], wearable devices [5, 8, 9, 17, 18, 22, 24, 27], etc. They can also be applied in the production of soundtracks for computer games [6, 10, 11]. The objectives underlying these contexts can be various: reflect the mood of game characters in the output of the system [6]; build recreational and therapeutic devices for people with neurological and physical disabilities [16, 25]; improve exercise performance [5, 17, 18, 22, 27]; adapt system output to a desired emotional state [1, 4, 7, 9, 10, 11, 12, 14, 15, 17, 21]; or, in a broad sense, dynamically express a desired emotion of/to an environment [3, 8, 18, 25, 26].

These systems are usually composed by a perception module that perceives the emotion from the environment by using data coming from different sources: psychophysiological sensors [1, 4, 5, 8, 9, 12, 15, 16, 17, 18, 27], cameras [3, 7, 13, 14, 17, 25], microphones [26], pressure sensors [26], temperature sensors [8], motion sensor [8, 13, 22, 25, 27], pollution sensors [8], physical data [4, 24] and self-report data [4, 9, 27]; and by an action module that reacts to the environment possibly by using different media (e.g., music and images). When using music to react to the environment, these systems usually make use of a composition/selection algorithm to generate music driven by emotional controls that intend to influence the emotional state of the user(s) of the environment. The algorithms used to generate music [3, 10, 11, 12, 21, 25, 26] deal with mapping between music and emotional data obtained with the referred sources. Musical features of the generated music reflect the intended emotional variations of the user(s).

Ada is a system especially relevant in this area [26]. It is an intelligent space that communicates moods, emotions and behaviors in real-time by using sound and light. Moods are defined by arousal and valence variables. Like others, Ada communicates by composing music to express the moods and emotions. Arousal variable changes tempo, volume and octave register. Valence variable changes consonance and pitch material. It is interesting the possibility to extend some of these systems to be used in a network of integrated music controllers [13]. These controllers create a collaborative interface between emotions and music generation.

3 EDME - Emotion-Based Music Engine

EDME is a system that produces music expressing a desired emotion. This objective is accomplished in 3 main stages: segmentation, selection and transformation; and 3 secondary stages: features extraction, sequencing and synthesis. We are using 2 auxiliary structures: a music base and a knowledge base. The music base has pre-composed MIDI music tagged with music features. It is prepared offiline for MIDIs of any musical style. The knowledge base is implemented as 2 regression models that consist of relations between each emotional dimension (valence and arousal) and music features. It can be adapted to the emotional feedback given by listeners (via questionnaires or psychophysiological signals). The characteristics of these auxiliary structures make the system easily adaptable to every context.

Aided by Figure 1 we will describe with more detail each of these stages. Pre-composed music of the music base is input to a segmentation module that produces fragments. These fragments must as much as possible be musically self-contained and express a single emotion. Segmentation discovers fragments by looking for note onsets with the higher weights. These weights are attributed according to the importance and degree of variation of five features: pitch, rhythm, silence, loudness and instrumentation. The features extraction module then obtains music features that are used to emotionally label the fragments with the help of the knowledge base, which are then stored in the music base.

The selection module obtains musical pieces with an emotional content similar to the desired emotion. These pieces are obtained from the music base, according to similarity metrics between desired emotion and music emotional content. Selected pieces can then be transformed to come even closer to the desired emotion. Transformation, also supported by the knowledge base, is applied in different characteristics of 5 groups of features: rhythm, melody, harmony, instrumentation and dynamics. The knowledge base has weights that control the degree of transformation for each characteristic. Sequencing module puts pieces produced by the transformation module in a smooth sequence by changing appropriate musical features. This sequence is given to a synthesis module that uses information about the General MIDI (GM) instruments [23] and timbral features to guide the selection of sounds from a library of sounds. This library is composed by various samples available from sample libraries of GM instruments (e.g., [28, 29]).

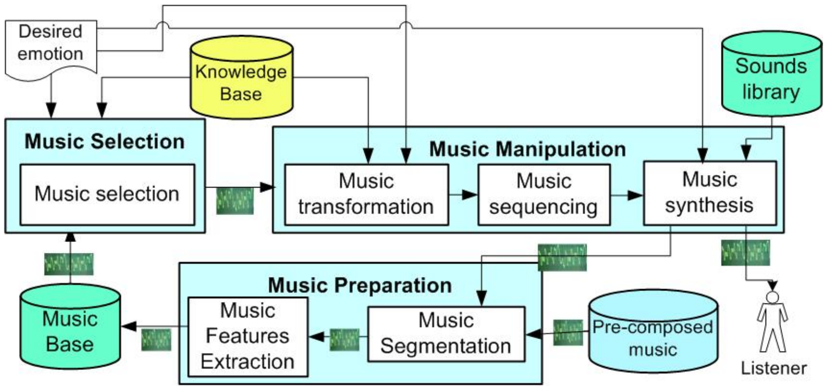

Fig. 1. Diagram portraying the 3 main stages (segmentation, selection and transformation) and 3 secondary stages (features extraction, sequencing and synthesis) of EDME.

4 Emotion-Based Interactive Installation

We developed an installation to provide an experimental context for assessing the interaction capabilities of EDME. We integrated the music engine in a multi-agent system built over JADE [2] that controls the overall behavior of the machinery. A computer connected to a camera, a projector, and speakers compose the setup (Figure 2).

The camera is placed on the ceiling of room. The interaction area representing by the grey circle is constrained by the field of view of the camera. One of the walls is a wide translucent screen where images are displayed under the computer's command.

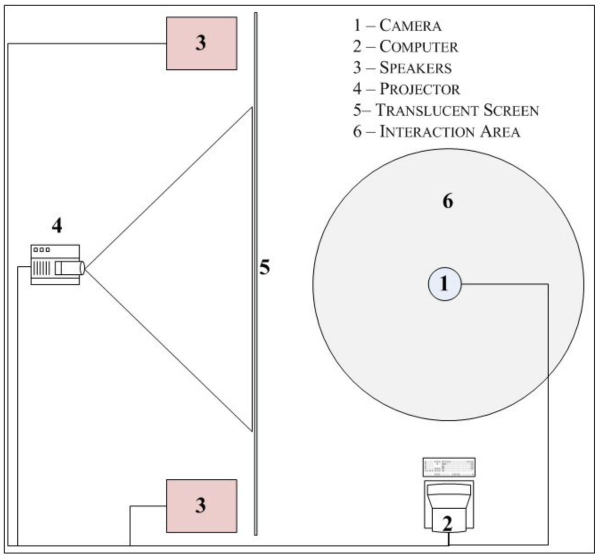

Fig. 2. Plan of the experimental setup.

The installation adapts the expressed emotions to the presence of people in the interaction area: the system expresses positive valence emotions on the presence of people and negative valence emotions when left alone. Moreover, people movement induces an increase in the arousal, whilst lack of activity induces a decrease. The system collects data from the environment through the camera and expresses emotions by means of the music it plays and the images it projects.

At the beginning of each trial session, the camera takes a picture of the environment. This first picture is considered to be the model. This image will represent the scenario of the experiment. As each session lasts for short time intervals, where the daylight and other environmental changes are not important, we assume that the scenario stays immutable during each session. Of course, if we needed to take a longer experience the light and environment changes could be prevented by a re-adaptation of the initial environment. Although the system prevents this, it is disabled for our experiences, since these experiences were made in closed rooms with artificial and constant light.

We are considering two environment variables: level of presence and movement. The system takes a new picture every second to update these variables. The level of presence is determined by computing the percentage of image change between the new picture and the scenario. This means that two persons induce a higher presence level than one. Moreover, the level is not dependent of people's movement, as long as they stay in the field of view of the camera. Movement is calculated by comparing two adjacent pictures. Using this method, only instant movement is calculated. As long as there is movement in the radius, the arousal increases, and when movement stops, arousal starts to decrease.

The environment information is then converted in a *stimulus* that is used to define the emotion that the system will express. Increase in the level of presence (e.g., one person enters the interaction area) produces a positive stimulus for valence; the reverse applies for decrease in the level of presence. As for movement, as long as there is activity in the interaction area, the stimulus for arousal increases. When movement stops, arousal starts to decrease. Arousal increase and decrease are linear with time.

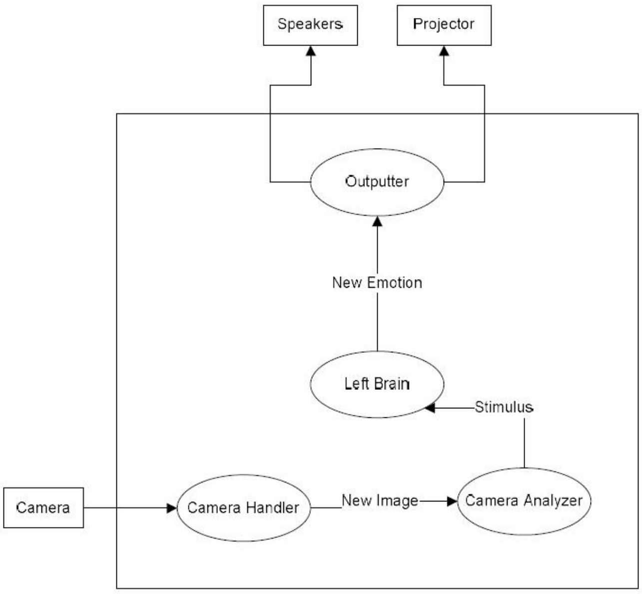

Fig. 3. Multi-agent system architecture.

The architecture of our multi-agent system is represented in Figure 3. An agent *Camera Handler* is in charge of gathering environment data. At the beginning of each trial session, the camera takes a picture of the environment. This image will represent the *scenario* of the experiment. As each session lasts for short time intervals, we assume invariance of daylight and other environmental conditions and, consequently, that the scenario stays immutable during the experiment.

An agent *Camera Analyser* converts the environment information in a *stimulus* that is used by an agent *Left Brain* to synthesise the new emotional state of the system according to the current emotional state and the received stimulus. Increase in the level of presence (e.g., one person enters the interaction area) produces a positive stimulus for valence; the reverse applies for decrease in the level of presence. As for movement, the stimulus for arousal increases as long as there is activity in the interaction area. When movement stops, arousal starts to decrease. Arousal changes are linear with time. Emotional states are synthesised with continuous change.

The new synthesised emotion is sent to the *Outputter* agent, who wraps EDME and is in charge of choosing the music to play and the image to project. The aim of the images is to reinforce the interaction experience provided by the installation. Images are chosen based in three levels of valence, sad, neutral and happy, and don't show arousal levels.

5 Experiments

We made two experiments with our system. The first one was a preliminary one that had the objective of observing the interaction between the system and participants (30 people); the second one was focused in obtaining participants feedback (via a questionnaire) about various components of the behavior of system. Figure 4 shows the two steps that composed both experiments. The first step consists in the perception of data from the environment: number of participants and quantity of movement. This data is used to define the emotional state of our system: valence (degree of happiness) and arousal (degree of activation). For instance, if it has the values 0 for arousal and 0 for valence, the system is very sad, if it has the values 10 for arousal and 10 for valence, the system is very happy and excited. The number of participants influence valence; the quantity of movement influence the arousal. The system's emotional state influences the way the system expresses to the environment by selecting different music and images to portray to the audience.

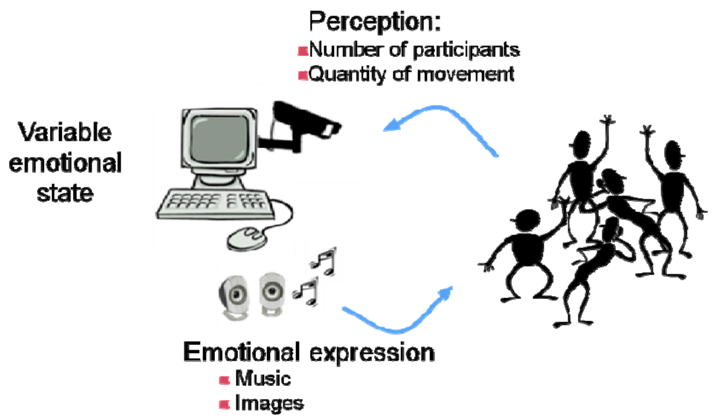

Fig. 4. Diagram illustrating the 2 steps of the two experiments: emotional perception and expression.

Participants interact with the system through sensors that detect presence and movement. At the core of the experiment, there is a computer that produces ambience music and projects images in a screen. Based on affective computing techniques, the computational system reflects in the musical features the variation of the number of presences and quantity of movement: for instance, produced music and images are happier with the presence of more people on the room; or can be more active when more movements are detected.

The first experiment was made in a session called "Discovering the Faculty on Saturdays". This session was attended by 30 students from secondary schools with ages from 10 to 18 years old. Figure 5 shows the environment with some participants.

Fig. 5. Picture taken during the first experiment.

In the second experiment we had groups of, approximately, 4 people interacting with the system in sessions of approximately 10 minutes. In the end of each session participants were asked to answer to a questionnaire with 7 questions (Table 1) with the objective to use obtained data/answers to evaluate the efficacy of the system. Each question was answered with values from the integer interval between 0 and 10. Sessions were attended by 23 people from the university (students and professors) with ages from 18 to 60 years old. Each session was conducted in three phases: first, the emotional state of the system was expressed by using only music; then, only images; and finally both media. This allowed us to evaluate the effect of each medium in emotional communication.

Table 1. 7 questions of the questionnaire given to the participants.

1)	The system expressed happiness with many presences and sadness with few presences
2)	The system expressed activation with much movement and relaxation with the lack of movement
3)	What is the importance of music in the emotional expression of the system
4)	What is the importance of images in the emotional expression of the system
5)	Music expressed expected emotion
6)	Images expressed expected emotion
7)	Efficacy of the system in the expression of the expected emotions

Both experiments were preceded by a series of preparatory steps: manual emotional classification of images and offline synthesis of music (Figure 6). Manual emotional classification consisted in attributing three levels of valence (sad, neutral, happy) based on personal background on this topic. Offline synthesis of music consisted in: segmentation of pre-composed music; automatically extraction of features from obtained segments; and automatic selection of music expressing different emotions. This music was then given to a synthesis module that converted selected MIDI music into mp3 audio music.

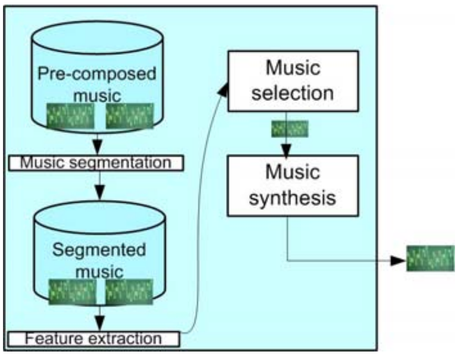

Fig. 6. Diagram illustrating the processes involved in the offline synthesis of music.

6 Results

The verbal feedback given by the participants of the first experiment was positive: the system interacted in an expected emotional way. This section is devoted to the analysis of the answers obtained with the questionnaire (Table 1) given to the 23 participants of the second experiment. Figure 7 presents the mean and standard deviation for the 7 questions of the questionnaire. We obtained high standard deviations when analyzing the importance of the image in the system and in the analysis of the expression of the expected emotion with the image. After the analysis of all the answers we came to the conclusion that, generally, the system correctly related arousal with the amount of movement and valence with the number of presences. Despite the fact that music seems a medium more important than images to express emotions, it was less successful than images in expressing the desired emotion. In general, the system was efficient in the transmission of the expected emotions. These conclusions give a first clue about the behavior of the system; however their significance is limited by the low number of participants (23), as well as, by the presence of some questions with multiple components (e.g., first and second questions).

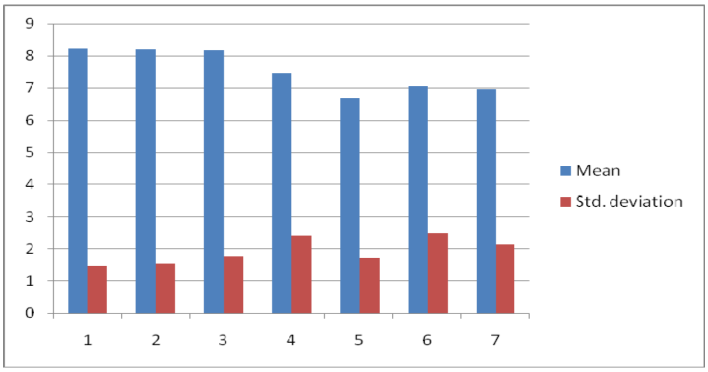

Fig. 7. Mean and standard deviations for the answers of the questionnaire.

Table 2. Correlation coefficients between the answers of the questionnaire (Table 1).

Answers	1)	2)	3)	4)	5)	6)	7)
1)	-	46	39	42	54	31	50
2)	46	-	29	45	56	44	**71**
3)	39	29	-	**82**	61	**67**	59
4)	42	45	82	-	55	**68**	60
5)	54	56	61	55	-	**71**	**76**
6)	31	44	67	68	71	-	**78**
7)	50	71	59	60	76	78	-
Mean	44	49	56	59	62	60	66

We also analyzed the correlation coefficients between the answers (Table 2) and concluded that the expression of the arousal of the system, the expression of expected emotions with music and images (answers 2, 5 and 6) has a high degree of correlation with the efficacy demonstrated by the system (answer 7). This leads us to have a special attention devoted in the analysis of the emotional effect of music and images, namely in the arousal dimension. The importance of music and images in the emotional expression and the expression of expected emotion in music (answers 3, 4 and 5) are in some way correlated with expression of expected emotion in images (answer 6). The importance of the emotional expression with music (answer 3) is highly correlated with the importance of the emotional expression with images (answer 4), which seems to reinforce the importance of using both media to express the emotional state of the system.

7 Conclusions

We prepared an installation that allows participants to experience and influence the emotional behavior of an interactive computer system. The experiences conducted show that people ascribe emotions to the system in a natural way. Results of the second experiment show the importance of both media (music and images) in the expression of the emotional state of the system; and the importance of analysing the emotional effect of music and images, namely in the arousal dimension. Our system can be, hopefully, more efficient by improving the EDME engine; by making an engine similar to EDME for images; and by complementing our work with studies of psychology.

We also obtained several suggestions from the participants to improve the system: synchronize music and images; use more emotional states for images; generate music and images in real-time; improve selection of music and images; decrease the delay of adaption of music to the environment; improve transitions among music; and have more sensors for movement and presences.

References

1. Arslan, B., Brouse, A., Simon, C., Lehembre, R., Castet, J., Filatriau, J. and Noirhomme, Q.: A real time music synthesis environment driven with biological signals. International Conference on Acoustics, Speech, and Signal Processing (2006).
2. Bellifemine, F., Poggi, A. and Rimassa, G.: JADE – A FIPA-compliant agent framework. Proceedings of PAAM, vol. 99, pp. 97-108 (1999).
3. Casella, P. and Paiva, A.: Magenta: An architecture for real time automatic composition of background music. International Workshop on Intelligent Virtual Agents. Springer, pp. 224–232 (2001).
4. Chung, J. and Vercoe, G.: The affective remixer: Personalized music arranging. Conference on Human Factors in Computing Systems. ACM Press New York, pp. 393–398 (2006).
5. Dornbush, S., Fisher, K., McKay, K., Prikhodko, A. and Segall, Z.: Xpod: A human activity and emotion aware mobile music player. International Conference on Mobile Technology, Applications and Systems (2005).
6. Eladhari, M., Nieuwdorp, R. and Fridenfalk, M.: The soundtrack of your mind: Mind music-adaptive audio for game characters. International conference on advances in computer entertainment technology. ACM Press NY (2006).
7. Funk, M., Kuwabara, K. and Lyons, M.: Sonification of facial actions for musical expression. New Interfaces for Musical Expression (NIME). National University of Singapore, pp. 127–131 (2005).
8. Gaye, L., Mazé, R. and Holmquist, L.: Sonic city: The urban environment as a musical interface. New Interfaces For Musical Expression (NIME). National University of Singapore, pp. 109–115 (2003).
9. Healey, J., Picard, R. and Dabek, F.: A new affect-perceiving interface and its application to personalized music selection. Workshop Perceptual User Interfaces (1998).
10. Hoeberechts, M., Demopoulos, R.J. and Katchabaw, M.: A Flexible Music Composition Engine. Audio Mostly 2007: The Second Conference on Interaction with Sound (2007).
11. Ka-Hing, J., Sze-Tsan, C., Kwok-Fung, C. and Chi-Fai, H.: Emotion-driven automatic music arrangement. International Conference on Computer Graphics and Interactive Techniques. ACM Press NY (2006).
12. Kim, S. and André, E.: Composing affective music with a generate and sense approach. Proceedings of Flairs 2004 - Special Track on AI and Music. AAAI Press (2004).

13. Knapp, R. and Cook, P.: Creating a network of integral music controllers. New Interfaces For Musical Expression, pp. 124–128 (2006).
14. Lyons, M. and Tetsutani, N.: Facing the music: a facial action controlled musical interface. Conference on Human Factors in Computing Systems. ACM Press NY, pp. 309–310 (2001).
15. McCaig, G. and Fels, S.: Playing on heart-strings: experiences with the 2hearts system. New interfaces for musical expression, pp. 1–6 (2002).
16. Miranda, E. and Boskamp, B.: Steering generative rules with the eeg: An approach to brain-computer music interfacing. Sound and Music Computing (2005).
17. Nakra, T.: Inside the Conductors Jacket: Analysis, Interpretation and Musical Synthesis of Expressive Gesture. PhD thesis, Massachusetts Institute of Technology (1999).
18. Oliver, N. and Flores-Mangas, F.: Mptrain: a mobile, music and physiology-based personal trainer. Conference On Human-Computer Interaction With Mobile Devices And Services. ACM Press NY, vol. 8, pp. 21–28 (2006).
19. Pantic, M., Sebe, N., Cohn, J. and Huang, T.: Affective multimodal human-computer interaction. ACM International Conference on Multimedia. ACM NY, pp. 669-676 (2005)
20. Picard, R.: Affective Computing. MIT Press Cambridge (1997).
21. Robertson, J., De Quincey, A., Stapleford, T. and Wiggins, G.: Real-time music generation for a virtual environment. Workshop on AI/Alife and Entertainment (1998).
22. Sakata, N., Kurata, T., Kourogi, M. and Kuzuoka, H.: Situated music: An application to interactive jogging. International Symposium on Wearable Computers (2006).
23. Selfridge-Field, E.: Beyond MIDI: the handbook of musical codes. MIT Press (1997).
24. Schiemer, G. and Havryliv, M.: Wearable firmware: The singing jacket. Proceedings of ACMA, pp. 66–71 (2004).
25. Wallis, I., Ingalls, T. and Campana, E.: Computer-Generating Emotional Music: The Design of An Affective Music Algorithm. International Conference on Digital Audio Effects (2008).
26. Wassermann, K., Eng, K., Verschure, P and Manzolli, J.: Live soundscape composition based on synthetic emotions. IEEE Multimedia, vol. 10, pp. 82-90 (2003).
27. Wijnalda, G., Pauws, S., Vignoli, F. and Stuckenschmidt, H.: A personalized music system for motivation in sport performance. IEEE Pervasive Computing, vol. 4, pp. 26–32 (2005).
28. Project SAM Symphobia: (accessed on July 14th 2009) http://www.projectsam.com/Products/Symphobia/
29. Garritan Personal Orchestra: (accessed on July 14th 2009) http://www.garritan.com/GPO-features.html