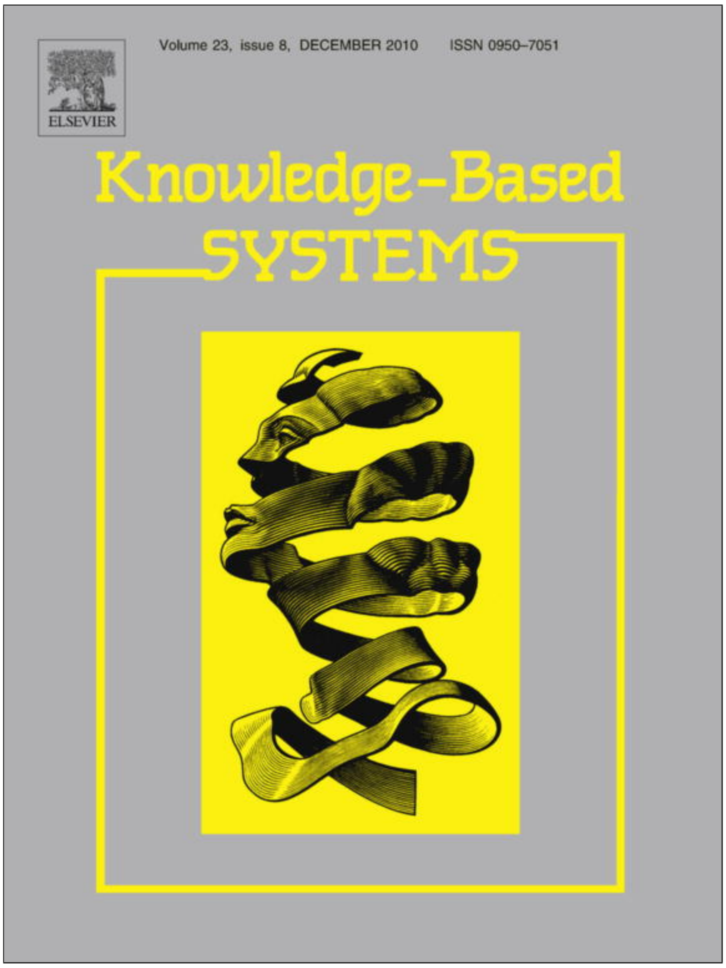
Volume 23, issue 8, DECEMBER 2010 ISSN 0950-7051
ELSEVIER
Knowledge-Based
SYSTEMS

Knowledge-Based Systems 23 (2010) 901–913

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A musical system for emotional expression

António Pedro Oliveira *, Amílcar Cardoso

Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Pólo II, Pinhal de Marrocos, 3030-290 Coimbra, Portugal

ARTICLE INFO

Article history:
Received 9 March 2010
Received in revised form 10 May 2010
Accepted 14 July 2010
Available online 18 July 2010

Keywords:
Knowledge-based system
Automatic music production
Expression of emotions
Music and emotions
Real-time system

ABSTRACT

The automatic control of emotional expression in music is a challenge that is far from being solved. This paper describes research conducted with the aim of developing a system with such capabilities. The system works with standard MIDI files and develops in two stages: the first offline, the second online. In the first stage, MIDI files are partitioned in segments with uniform emotional content. These are subjected to a process of features extraction, then classified according to emotional values of valence and arousal and stored in a music base. In the second stage, segments are selected and transformed according to the desired emotion and then arranged in song-like structures.

The system is using a knowledge base, grounded on empirical results of works of Music Psychology that was refined with data obtained with questionnaires; we also plan to use data obtained with other methods of emotional recognition in a near future. For the experimental setups, we prepared web-based questionnaires with musical segments of different emotional content. Each subject classified each segment after listening to it, with values for valence and arousal. The modularity, adaptability and flexibility of our system's architecture make it applicable in various contexts like video-games, theater, films and healthcare contexts.

1. Introduction

The scientific challenge of automatically producing music with an appropriate emotional content has involved much research in emotional and musical domains. Throughout the history, many scientists have studied emotions [8,12,14,22,37]; however, there is no consensus in their definition [41]. We accept emotions as corresponding to the manifestation of our psychophysiological state [20]. Music is a powerful stimulus capable of influencing our emotions. This has been proved by research findings on its perception and expression [9,23,29,47,52]; and more recently by studies that have found relations between musical features and emotions [16,18]. For instance, tempo is widely accepted as having direct influence on the pleasantness of emotions.

Systems that effectively produce music expressing specific emotions are relevant to be used in contexts where there is a need to create environments capable of inducing certain emotional experiences. The production of soundtracks for video-games, films and theatre are examples. They can also be applied in hospitals, shopping centres, gymnasiums and houses of worship places. This motivated the development of Emotion-Driven Music Engine (EDME), a system with the mentioned capabilities.

* Corresponding author. Tel.: +351 962338424.
E-mail addresses: apsimoes@dei.uc.pt, apsimoes@student.dei.uc.pt (A.P. Oliveira), amilcar@dei.uc.pt (A. Cardoso).

0950-7051/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2010.06.006

2. Related work

Many important contributions to our work derived from research done in areas of Computer Science, Psychology and Music. Concerning the representation of emotions, the prevailing alternative is between discrete and dimensional systems with two- or three dimensions [7]. The most common interpretation for dimensions construes them as: arousal (activation/relaxation), valence (pleasantness/unpleasantness) and dominance (degree of control over the emotional state). The first two dimensions capture most of the empirical variance, which explains that the third one is often ignored.

The main source of knowledge for systems like EDME is empirical data that relates emotions and musical features [13,16,18,24,25,42,50]. Livingstone et al. [25] distinguishes the perceived emotion and the experienced emotion. We are currently focusing our research on the effect of structural and performative features on the perceived emotions. In the long term, we are also interested on the effects of these features on the experienced emotions [40,43]. After an extensive review of empirical data available on the literature, we made a systematization of the relevant features that are common to four types of music: happy, sad, activating and relaxing (Fig. 1).

These scientific advances have been the key source of inspiration to four main approaches being used to tackle the scientific challenge of our work. The first approach consists in composing/arranging music, e.g., by generating music from scratch according

Musical Feature		Happy music	Sad music	Activating music	Relaxing music
Instruments	timbre	piano, strings instruments, few harmonics, bright, percussion instruments	timpani, violin, woodwind instruments, few harmonics, dull, harsh	brass, low register instruments, timpani, harsh, bright, percussion instruments	woodwind instruments, few harmonics, soft
Dynamics	loudness	high	low	high	low
	articulation	staccato	legato	staccato	legato
	articulation variab.	large	small	-	-
	sound variability	low	-	-	-
Rhythm	tempo	fast	slow	fast	slow
	note density	high	low	high	low
	note duration	small	large	small	large
	tempo variability	small	-	-	-
	duration contrast	sharp	soft	-	-
Melody	pitch register	high	low	-	-
	pitch repetition	high	low	high	low
	stable/ expect notes	accented	-	-	-
	unstab/ unexp notes	accented	-	-	-
Harmony	harmony	consonant	dissonant	complex, dissonant	-
	scale	major, pentatonic	minor, diminished	-	-

Fig. 1. Features of happy, sad, activating and relaxing music.

to emotional cues [46,50]. The lack of flexibility in adapting the musical output to different styles set automatic music composition outside our aims. The second approach consists in selecting pre-composed music. It requires the extraction of musical features – statistical and perceptual – that are subsequently used to make recommendation/classification models [1,48,56,57]. The third approach consists in transforming/adapting pre-composed music – currently, this approach is only viable if working at a symbolic representation level. This can be done through a knowledge-based control of structural factors of pre-composed musical scores [25,53,54]. These two last approaches produce solutions with low quality when the emotional content of the source music is far from the required one. The sequential use of classification stage before the transformation overcomes the limitation of both approaches. This drives us to the fourth approach that consists in combining some of the above-mentioned alternatives. Chung and Vercoe [6], for example, uses mixed techniques, but this work used an approach that seems quite ad-hoc and no technical details are available.

3. Emotion-Driven Music Engine (EDME)

From the analysis described in the previous section, we found four opportunities to contribute to the advance of the state-of-the-art: the representation of emotions in both the discrete and dimensional spaces; the systematization of the relations between emotions and musical features in a knowledge base; the development of algorithms that control the emotional content of music; and the development of a parameterizable architecture, suited to the prosecution of experimental work. To accomplish these aims we designed EDME, a flexible and adaptable system composed by four main modules (segmentation, classification, selection and transformation) that control the emotional content of music; three secondary modules (features extraction, sequencing and synthesis) responsible for doing work necessary for the main modules; four auxiliary structures (music base, knowledge base, pattern base and a library of sounds) that store content useful for some of the modules; and an user interface.

The system works in two stages, one offline and another online. In the offline stage (Fig. 2), the segmentation module generates musical segments that express only one emotion by analyzing features of pre-composed music. These segments are given to the module of features extraction that obtains features used by the classification module. This module uses a knowledge base to label the segments with emotional values of valence and arousal. MIDI music emotionally classified is then stored in a music base.

In the online stage (Fig. 3), the selection module calculates the distance between these values and desired emotions; then, it selects from a music base segments with the minimum distances.

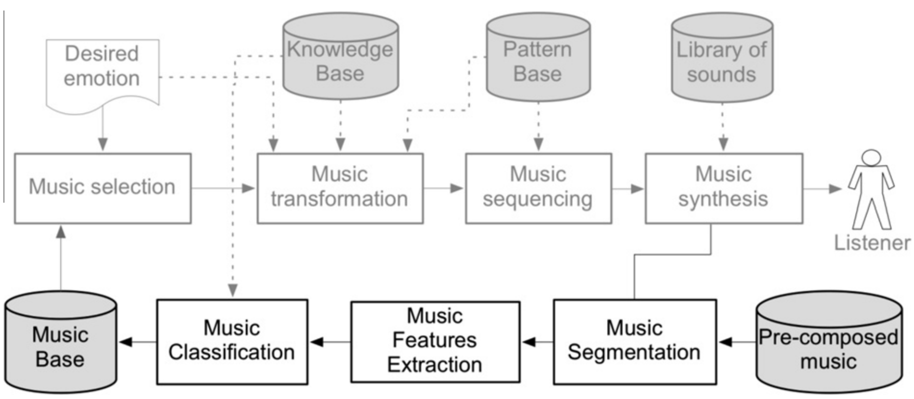

Fig. 2. EDME architecture: offline stage.

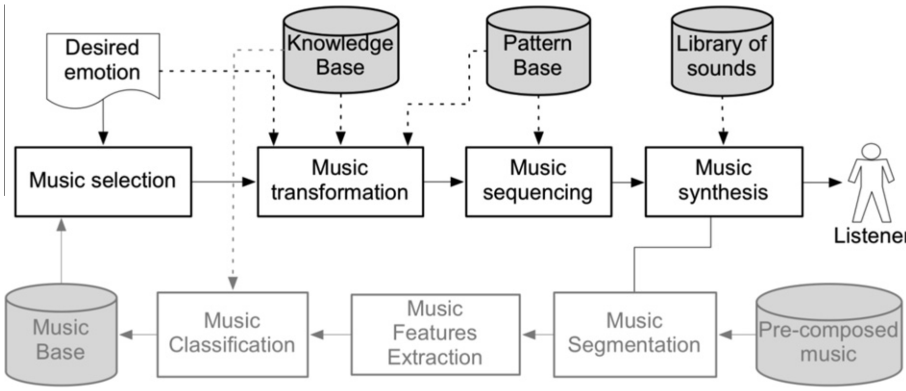

Fig. 3. EDME architecture: online stage.

The transformation module approximates the emotional content of the segments to the desired emotion by changing features emotionally relevant. The sequencer module packs the segments to form songs and the synthesizer deals with the selection of sounds to convert the MIDI output into audio. The input of the system, a desired emotion, is defined from a list of discrete emotions or from a bi-dimensional emotional space. This input is controlled with a user interface.

3.1. Music segmentation

The pre-composed music consists of standard MIDI files compiled from websites or other sources, or possibly composed on purpose. These files are polyphonic and can be of any musical style. The segmentation module uses these files to produce segments as much as possible with a musical sense of its own and expressing a single emotion (Fig. 4). We believe that obtaining smaller musical pieces decreases the probability of finding more than one emotion in each segment. We made some perception tests with three segmentation algorithms available on the MIDI Toolbox [11] and found that Local Boundary Detection Model [4] obtained the best results.

This module starts by attributing weights to each note onset by using an adaptation of Local Boundary Detection Model (LBDM), a rule-based model based on gestalt principles of change and proximity. These weights are attributed according to the musical importance, degree of proximity and degree of variation of five features: pitch, rhythm, silence, loudness and instrumentation. The degree of proximity and the degree of variation are calculated according to the LBDM; musical importance is a parameter that was defined after making some perception tests with the aim of finding the best points of segmentation. The module searches for plausible points of segmentation according to the weights attributed at each note onset. There is a threshold defined to reduce the weights' search space: note onsets with weights below this threshold are not considered. The length of obtained segments is defined by a minimum (min) and maximum (max) number of bars. The module starts by searching for a plausible point of segmentation that corresponds to the maximum weight obtained between the first bar of music file + min and the first bar + max. This process is then iterated, starting from the bar of the last point of segmentation, till the end of the file is reached.

3.2. Music features extraction

The features extraction module uses toolboxes that obtain features supposed to be relevant to our system according to the literature (e.g., [16,18]. The JSymbolic [27], MIDI Toolbox [11] and JMusic [44] extract symbolic features; MIR Toolbox [21] and Psysound Toolbox [3] extract audio features. We also developed our own algorithms to extract additional symbolic and audio features (e.g., average loudness and spectral similarity). Average loudness corresponds to the average velocity of all the MIDI notes. Spectral similarity calculates a similarity matrix with the help of MIR Toolbox [21] in order to find the difference between consecutive frames of the frequency spectrum. It reflects the smoothness of the music (the changes of features along the music). Both have a relation with the arousal of music [42]. At this moment, it is possible to extract 476 features that belong to five groups: instrumentation, dynamics, rhythm, melody and harmony (see Fig. 5).

This module labels each segment with emotionally relevant musical features. The relevance of each feature is defined according to empirical results obtained both from the literature (e.g., [16,25] and from our experiments (see Section 4).

3.3. Music classification

EDME has a knowledge base composed by two normal regression models [51] – one for each emotional dimension. The models provide weighted relations between music features and the dimension in question. They were built by applying feature selection and regression algorithms [55] on experimental data (see Section 4 for more details).

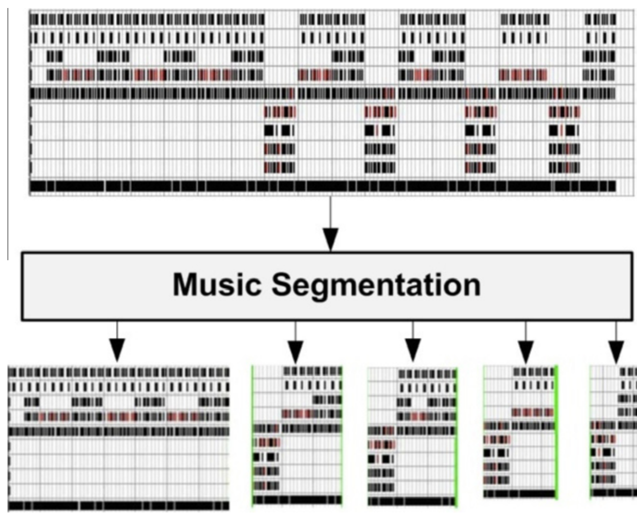

Fig. 4. Input and output to the segmentation module.

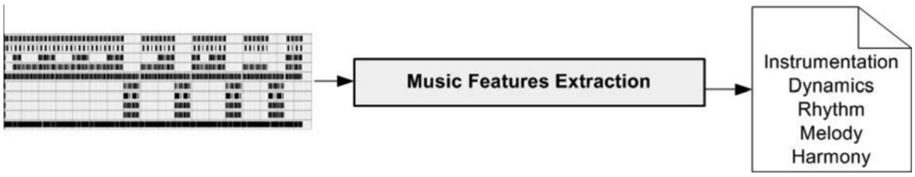

Fig. 5. Input and output of the module of features extraction.

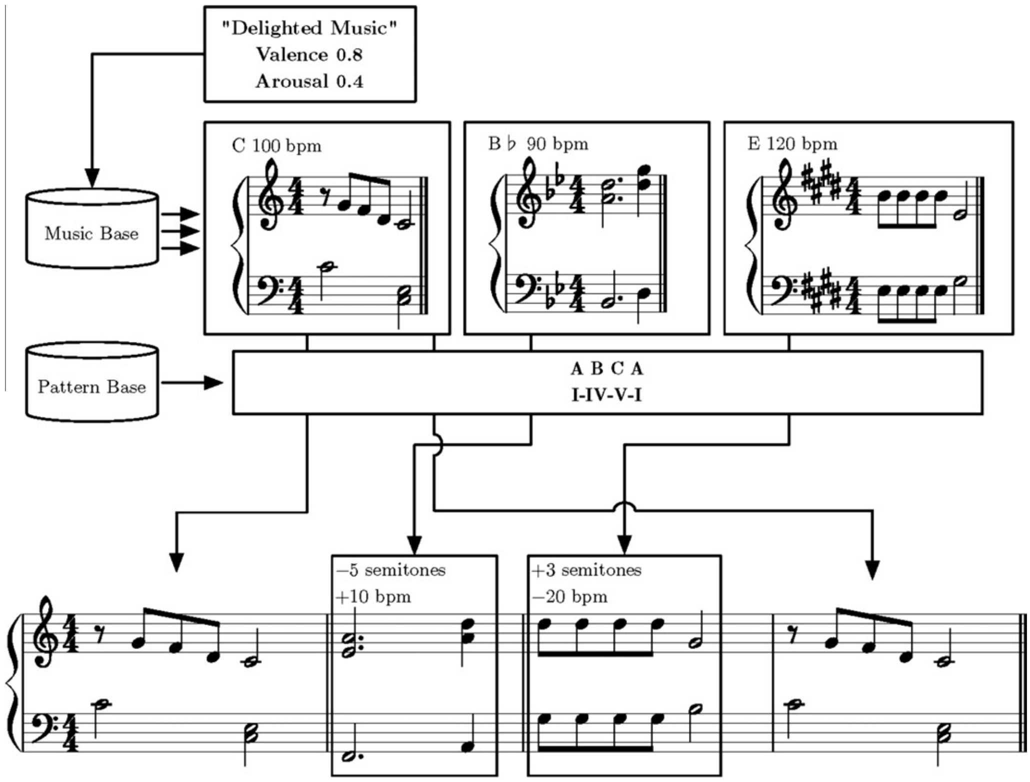

Fig. 6. Sequencing example.

The classification module uses the knowledge base to determine the emotional content of a segment by computing the weighted sum of the values of the features obtained for each emotional dimension with the module of features extraction:

$$Valence = \sum_{i=0}^{n} valenceFeature\_Weight_i * valenceFeature\_Value_i \quad (1)$$

$$Arousal = \sum_{i=0}^{n} arousalFeature\_Weight_i * arousalFeature\_Value_i \quad (2)$$

The computed values are stored (as tags) with the segments in a music base.

3.4. Music selection

The selection module compares the emotional content of each segment to the desired emotion using the Euclidean distance. The results of the various comparisons are used to put the segments in a list ordered by the degree of similarity to the desired emotion. This module retrieves the segments that are on the top of the list. The number of segments that are retrieved is customizable.

3.5. Music transformation

The transformation module also uses the knowledge base to approximate the emotional content of selected segments to the desired emotion. This module starts by calculating two distances using the Euclidean metric: the distance between the valence of each selected segment and the valence of the desired emotion; and the distance between the arousal of each selected segment and the arousal of the desired emotion. In order to minimize both distances, the module transforms musical features by a specific quantity. This quantity depends on the quotient between each distance and the weight of the feature defined in the regression models for each emotional dimension. We developed six algorithms that transform tempo, pitch register, musical scale, instruments, articulation and the contrast of the duration of notes.

Let us give an example. Suppose we want a desired emotion of *Valence*, *Arousal* = (0.95, 0.4) with *Valence*, *Arousal* $\in [-1, 1]$ and the segment with the closest emotional content has an emotion of *Valence*, *Arousal* = (0.5, 0.4). The dimension of arousal does not need to be changed; however, the system needs to change the dimension of valence from 0.5 to 0.95. If the regression model of valence has an equation of 0.005 $*$ tempo + 0.05 $*$ pitch and the retrieved segment has a tempo of 50 and a pitch of 50, in order to meet the desired emotion the system may transform the tempo to 120 and the pitch to 70.

3.6. Music sequencing

The sequencing module resorts to a pattern base to pack the segments to form a sequence of songs. Each pattern defines a song structure and the harmonic relations between the segments of the structure (e.g., popular song patterns like AABA). Segments are arranged in order to match the tempo and pitch of the pattern. The tempo of the segments is normalized to their average tempo. The pitch is raised or lowered, by comparing the key of the current pattern with the key of the non-transformed segments.

We present an example in Fig. 6 where the user wants to hear music expressing a delighted emotion, represented as *Valence*, *Arousal* = (0.8, 0.4). The system selects three MIDI segments (the ones closer to the desired emotion) to match the current – ABCA – pattern. The first segment, with C as the tonic and a tempo of 100 bpm, acts as the root of the pattern. The second segment needs transformations to match the tempo (+10 bpm) and the pitch (the IV-subdominant of C is F, so −5 semitones gets B to F). The third segment needs transformations to match the tempo (−20 bpm) and the pitch (the V-dominant of C is G, so +3 semitones gets E to G). Finally the first segment is repeated to end the pattern.

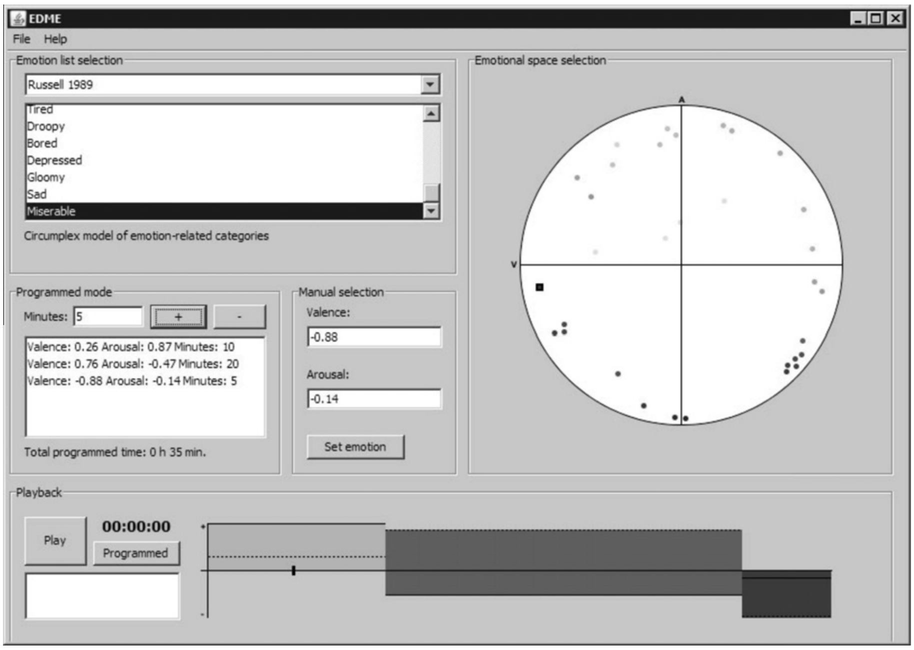

Fig. 7. User interface of EDME.

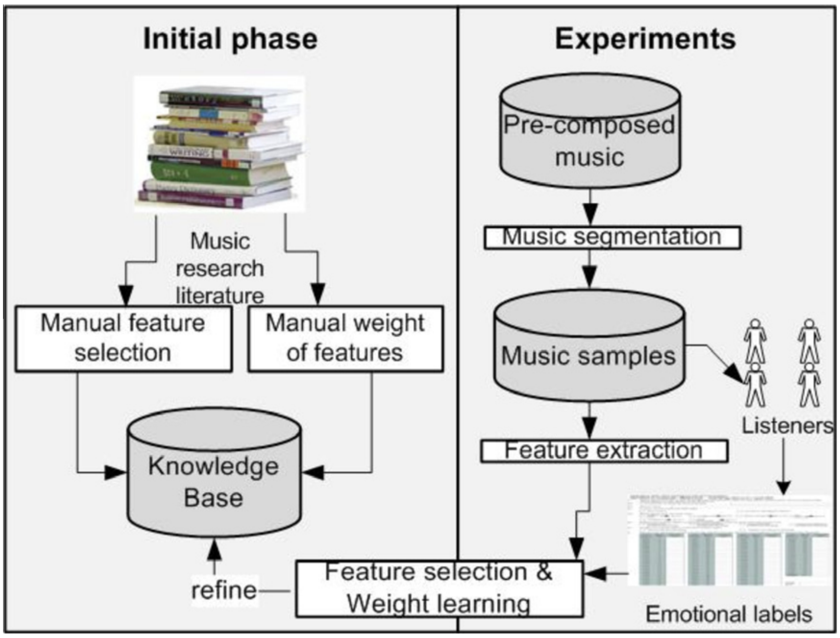

Fig. 8. Stages of the experiments.

Table 1
Results of 10-fold cross-validation for valence and arousal – first experiment.

Emotional dimension	CC	MAE	RMSE	Best features	Weight
Valence	0.76	0.75	0.91	Average time between attacks	−0.50
				Variability of note duration	−0.55
Arousal	0.77	0.86	1.06	Average note duration	−0.48
				Average time between attacks	−0.35
				Importance of high register	−0.45
				Note density	0.09

Table 2
Results of 10-fold cross-validation for valence and arousal – second experiment.

Emotional dimension	CC	MAE	RMSE	Best features	Weight
Valence	0.70	0.85	1.04	Average note duration	+0.31
				Initial tempo	−0.48
				Key mode	−0.18
				Note density	+0.34
Arousal	0.77	0.84	1.04	Average note duration	−0.84
				Initial tempo	0.42
				Note density	0.41

Table 3
Best audio features for valence and arousal – second experiment.

Emotional dimension	Best features
Valence	Spectral sharpness (Amber)
	Spectral sharpness (Zwickler)
	Timbral width
	Spectral loudness
Arousal	Spectral sharpness (Amber)
	Spectral dissonance (Sethares)
	Spectral sharpness (Zwickler)
	Spectral similarity

The segments are sequenced in order to be perceived as a single part with distinct harmonic relations and equal tempo.

3.7. Music synthesis

EDME has a library of sounds composed by samples for the instruments of General MIDI 1 standard. These samples were obtained from [17,38], and a personal library of [45]. The synthesis module calculates the emotional content of the samples of each instrument according to the spectral dissonance and spectral sharpness. Dissonance is used to label arousal and sharpness is used to label valence [33]. The module is using Psysound toolbox [3] to extract these audio features. The emotional content drives the selection of sounds from the library in order to produce an audio output.

3.8. User interface

The system can be controlled in real-time through a user interface (represented in Fig. 7[1]) or be driven by an external system providing an emotional specification [26]. The input specifies values of valence and arousal. While playing, EDME responds to input changes by quickly adapting the music to a new user-defined emotion.

The user interface serves the purpose of letting the user choose the desired emotion in different ways. It is possible for the user to directly type the values of valence and arousal the music should have. Other way is through a list of discrete emotion the user can choose from. It is possible to load several lists of words denoting emotions to fit different uses of the system. For example, Ekman [12] has a list of generally accepted basic emotions. Russell [39] and Mehrabian [28] both have lists which map specific emotions to dimensional values (using two- or three dimensions). Juslin and Laukka [19] propose a specific list for emotions expressed by music. Another way to choose the emotional state of the music is through a graphic representation of the valence-arousal emotional space, based on FeelTrace [5]: a circular space with valence dimension is in the horizontal axis and the arousal dimension in the vertical axis.

4. Experiments

The heart of the system is the classification module, where EDME establishes a bridge between the emotional and the musical dimensions. The overall aim of the realized experiments was to improve the structures that support this module: the regression models of the knowledge base. These models are independent from social variables like the age of the listeners and musical variables like the musical style. At an initial phase, a first version of the knowledge base was manually built [30] by considering empirical data collected from works of Music Psychology [9,16,42]. Positive/negative weights were defined for each feature according to their influence on the emotional dimensions (Fig. 1). Then, three experiments [31–33,35] were conducted to build regression models and to successively refine their set of features and corresponding weights with data obtained from web-based questionnaires[2,3,4]. The third experiment also aimed to verify the effectiveness of the regression models in supporting the transformation module.

Fig. 8 presents an overview of different stages of the experiments. They started with the segmentation of MIDI music to obtain segments that could express only one kind of emotion. Then, feature extraction algorithms of third party software [3,11,21,27,44] were applied to label the segments with music features. The relations of the knowledge base were used to label music with emotional content. For each experiment, we prepared a set of segments with different emotional content and made them available in the

[1] http://www.youtube.com/watch?v=xFbkPlQJ1WQ.

[2] http://student.dei.uc.pt/~apsimoes/PhD/Music/icmc08/.
[3] http://student.dei.uc.pt/~apsimoes/PhD/Music/smc08/.
[4] http://student.dei.uc.pt/~apsimoes/PhD/Music/smc09/.

Table 4
Results of 10-fold cross-validation for valence and arousal – third experiment.

Emotional dimension	CC	MAE	RMSE	Best features	Weight
Valence	0.69	0.76	0.97	Average time between attacks	–0.18
				Num. of unpitched instruments	+0.20
				Overall dynamic range	+0.32
				Percussion prevalence	–0.12
				Variability of note duration	–0.31
Arousal	0.71	0.81	0.99	Note density	0.29
				Percussion prevalence	0.12
				Variability unpitched instruments	0.15

Table 5
Results of 10-fold cross-validation for valence and arousal considering only the melodic line.

Emotional dimension	CC	MAE	RMSE	Best features	Weight
Valence – data of first experiment	0.79	0.61	0.87	Average note duration	–0.04
				Rhythmic variability	–0.35
				Staccato incidence	+0.18
				Time prevalence of koto	+0.22
				Variability of note duration	–0.50
Valence – data of second experiment	0.62	0.94	1.18	Average time between attacks	–0.47
				Initial tempo	+0.59
				Maximum note duration	–0.06
				Variability of note duration	+0.04
				Variation of dynamics	+0.25
Valence – data of third experiment	0.41	1.00	1.25	Average note duration	–0.16
				Comb. streng. two strong. pulses	–0.25
				Minimum note duration	–0.32
				Strength strong. rhythmic pulse	+0.11
				Average note duration	–0.48
Arousal – data of first experiment	0.85	0.64	0.85	Initial tempo	0.29
				Maximum note duration	–0.25
				Most common pitch prevalence	0.29
Arousal – data of second experiment	0.72	0.89	1.14	Average note duration	–0.09
				Average time between attacks	–0.83
				Initial tempo	0.41
				Variation of dynamics	0.43
Arousal – data of third experiment	0.54	0.94	1.20	Number of common pitches	0.05
				Rel. streng. common mel. interval	0.14
				Variation of dynamics	0.40

Fig. 9. Scatterplot of emotional data of first, second and third experiments.

Table 6
Results of 10-fold cross-validation for valence and arousal.

Emotional dimension	CC	MAE	RMSE	Best features	Weight
Valence	0.59	0.92	1.14	Average note duration	–0.32
				Average time between attacks	–0.38
				Initial tempo	+0.34
				Key mode	–0.08
Arousal	0.74	0.87	1.07	Average note duration	–0.11
				Initial tempo	+0.38
				Note density	+0.49
				Percussion prevalence	+0.21

Classifier	GP	IR	LMS	LR	MP	PR	RBF	SLR	SMO	IBK	KS	LWL	AR	BAG	ES	RSS	RD	CR	DT	M5R	DS	M5P	REP	Measure
First experiment cross-validation	0.69	0.66	**0.72**	**0.73**	0.63	-	0.19	0.55	**0.76**	0.54	0.68	0.49	**0.71**	0.60	0.61	0.59	**0.72**	0.45	0.39	0.69	0.42	**0.73**	0.52	CC
	0.89	**0.83**	**0.83**	**0.78**	0.97	-	1.31	1.00	**0.75**	1.15	0.98	1.04	0.87	0.99	0.95	1.00	**0.83**	1.07	1.11	0.84	1.12	**0.79**	1.00	MAE
	1.07	1.07	**0.97**	**0.95**	1.15	-	1.41	1.19	**0.91**	1.27	1.12	1.27	1.11	1.11	1.11	1.17	**0.99**	1.27	1.30	1.01	1.37	**0.95**	1.21	RMSE
First experiment training/test split	**0.81**	0.46	**0.85**	0.54	0.67	-	0.70	0.20	0.79	0.53	**0.88**	**0.91**	**0.86**	**0.81**	0.70	0	**0.90**	0	0	0.38	0.70	0.37	0	CC
	0.96	1.45	1.30	0.93	1.41	-	**0.79**	1.62	**0.84**	**0.75**	1.14	0.96	1.06	1.29	0.92	1.54	1.13	1.12	1.43	1.37	**0.85**	1.33	1.54	MAE
	1.03	1.54	1.41	**1.05**	1.50	-	**1.09**	1.82	**1.02**	**0.97**	1.20	1.12	1.17	1.34	**1.06**	1.61	1.18	1.21	1.51	1.65	1.09	1.60	1.61	RMSE
Second experiment cross-validation	**0.72**	0.59	**0.71**	**0.71**	0.68	**0.71**	0.53	0.61	**0.70**	0.57	0.56	0.50	0.61	0.63	0.61	0.63	0.50	0.39	0.53	**0.71**	0.34	**0.71**	0.55	CC
	0.84	0.98	**0.84**	**0.83**	0.87	**0.83**	1.01	0.98	0.85	1.09	0.98	1.04	0.95	0.93	0.95	0.93	1.10	1.12	1.08	**0.83**	1.14	**0.83**	0.97	MAE
	1.01	1.16	1.03	**1.02**	1.11	**1.01**	1.22	1.14	1.04	1.40	1.22	1.26	1.20	1.12	1.15	1.11	1.34	1.34	1.24	**1.02**	1.39	**1.02**	1.24	RMSE
Second experiment training/test split	**0.83**	0.66	**0.83**	**0.80**	0.63	**0.82**	0.59	0.72	**0.82**	0.61	0.56	0.66	0.73	0.77	0.59	0.74	0.71	0.48	0.64	**0.80**	0.54	**0.80**	0.64	CC
	0.91	1.04	**0.82**	**0.88**	1.06	0.95	1.14	1.06	**0.87**	1.10	1.08	1.04	0.90	0.98	1.15	1.01	0.91	1.13	1.08	**0.88**	1.09	0.88	1.06	MAE
	1.02	1.21	**0.96**	**1.00**	1.26	**0.97**	1.35	1.17	**1.00**	1.36	1.32	1.21	1.08	1.12	1.31	1.17	1.12	1.39	1.22	**1.00**	1.33	**1.00**	1.22	RMSE
Third experiment cross-validation	**0.66**	0.49	**0.64**	**0.68**	**0.66**	**0.68**	0.40	0.37	**0.69**	0.55	0.60	0.62	0.48	0.56	0.49	0.58	0.41	0.15	0.36	**0.65**	0.12	**0.68**	0.51	CC
	0.77	1.00	0.83	**0.77**	0.83	**0.78**	1.00	1.01	**0.76**	1.00	0.87	0.87	0.93	0.83	0.90	0.89	0.96	1.06	1.07	**0.80**	1.08	**0.78**	0.90	MAE
	1.00	1.20	**1.04**	**0.98**	1.08	**0.97**	1.23	1.28	**0.97**	1.27	1.13	**1.06**	1.28	1.10	1.20	1.09	1.31	1.40	1.33	**1.02**	1.44	**0.98**	1.16	RMSE
Third experiment training/test split	**0.71**	0.45	**0.75**	**0.75**	0.50	**0.76**	0.38	0.41	**0.77**	0.62	0.65	0.69	0.51	0.67	0.69	0.51	0.51	0	0.33	0.44	0.37	0.44	0	CC
	0.90	1.07	**0.87**	**0.77**	1.23	**0.75**	1.13	1.10	**0.78**	0.96	**0.87**	**0.87**	1.09	0.91	0.95	1.12	1.09	1.27	1.11	1.09	1.18	1.09	1.24	MAE
	1.08	1.25	**1.07**	**0.97**	1.60	**0.95**	1.30	1.28	**0.99**	1.18	1.11	**1.05**	1.26	1.11	1.15	1.31	1.32	1.39	1.35	1.26	1.31	1.26	1.42	RMSE
Overall performance	**0.74**	0.55	**0.75**	**0.70**	0.64	**0.74**	0.47	0.48	**0.76**	0.57	0.66	0.65	0.65	0.67	0.62	0.51	0.53	0.25	0.38	0.61	0.42	0.62	0.37	CC
	0.88	1.06	**0.92**	**0.83**	1.06	**0.83**	1.06	1.13	**0.81**	1.00	0.99	0.97	0.97	0.99	0.97	1.08	1.00	1.13	1.15	0.97	1.08	**0.95**	1.12	MAE
	1.04	1.24	**1.08**	**1.00**	1.28	**0.98**	1.27	1.31	**0.99**	1.24	1.18	1.16	1.18	1.15	1.16	1.24	1.21	1.33	1.33	1.16	1.32	**1.14**	1.32	RMSE

Fig. 10. Classifiers performance for valence. GP – Gaussian Process; IR – Isotonic Regression; LMS – Least Mean Square; LR – Linear Regression; MP – Multilayer Perceptron; PR – Pace Regression; RBF – Radial Basis Function; SLR – Simple Linear Regression; SMO – SVM Regression; IBK – Instance-Based K-Nearest Neighbor; KS – K Star; LWL – Locally-weighted Learning; AR – Additive Regression; BAG – Bagging; ES – Ensemble Selection; RSS – Random SubSpace; RD – Regression By Discretization; CR – Conjuctive Rule; DT – Decision Table; M5R – M5 Rules; DS – Decision Stump; M5P – M5 Trees; REP – REP Tree.

Classifier	GP	IR	LMS	LR	MP	PR	RBF	SLR	SMO	IBK	KS	LWL	AR	BAG	ES	RSS	RD	CR	DT	M5R	DS	M5P	REP	Measure
First experiment cross-validation	**0.82** **0.89** **1.13**	0.16 1.32 1.86	0.26 1.75 2.68	0.64 0.98 1.33	**0.80** **0.77** **1.09**	- - -	0.62 0.98 1.29	0.44 1.15 1.52	**0.77** **0.86** **1.06**	**0.75** **0.67** **1.11**	0.63 **0.85** 1.31	0.55 0.92 1.46	**0.76** **0.67** **1.09**	0.49 1.07 1.42	0.42 1.21 1.48	0.14 1.34 1.64	0.50 1.22 1.52	0.17 1.44 1.73	0.63 0.97 1.28	0.56 1.11 1.44	0.20 1.22 1.81	0.56 1.11 1.44	0.10 1.39 1.68	CC MAE RMSE
First experiment training/test split	**0.89** **0.95** **1.26**	0.38 1.07 1.80	**0.81** 1.09 **1.28**	**0.84** 1.02 **1.25**	**0.95** **0.86** **1.02**	- - -	**0.98** **0.25** **0.30**	0.30 1.32 1.93	**0.94** **0.95** **1.13**	0.40 1.26 1.77	0.21 1.45 1.99	0.38 1.15 1.73	0.37 1.15 1.90	0.61 **0.89** 1.47	0.39 **0.98** 1.63	0.54 1.05 1.51	0.24 1.32 1.79	0.52 **0.96** 1.56	0.44 1.11 1.61	**0.88** 0.99 **1.23**	0.52 **0.93** 1.61	**0.88** 0.99 **1.23**	0.52 **0.95** 1.64	CC MAE RMSE
Second experiment cross-validation	**0.81** **0.79** **0.96**	**0.80** **0.76** **0.97**	**0.79** **0.81** **1.02**	**0.77** **0.81** **1.02**	0.74 0.90 1.12	**0.77** 0.82 **1.03**	0.74 0.88 1.07	0.61 0.97 1.32	**0.77** 0.84 1.05	0.71 0.96 1.20	0.73 0.89 1.10	0.75 0.86 1.05	**0.80** **0.75** **0.96**	**0.79** **0.77** **0.97**	**0.80** **0.75** **0.96**	0.76 0.88 1.04	0.75 0.86 1.09	0.69 0.96 1.15	0.68 0.97 1.18	0.76 0.85 1.06	0.73 0.90 1.08	**0.77** 0.82 **1.03**	0.76 0.83 1.04	CC MAE RMSE
Second experiment training/test split	**0.85** 0.92 1.07	0.79 0.89 1.08	0.81 **0.86** **1.04**	0.84 0.90 1.06	**0.86** 0.96 1.15	0.84 0.90 1.06	**0.85** **0.84** 1.05	0.76 1.12 1.27	**0.85** 0.89 **1.04**	0.79 0.89 1.08	0.83 0.87 1.05	0.81 0.88 1.08	0.81 **0.82** **1.04**	0.83 **0.85** **1.02**	0.79 0.91 1.09	0.79 0.94 1.11	**0.86** **0.74** **0.92**	0.79 0.92 1.13	0.57 1.20 1.45	**0.86** **0.85** **0.98**	0.79 0.90 1.11	**0.87** **0.85** **0.98**	0.78 0.93 1.12	CC MAE RMSE
Third experiment cross-validation	**0.71** **0.79** **1.00**	0.62 0.90 1.11	**0.68** 0.86 1.04	**0.70** **0.82** **1.00**	0.56 1.04 1.30	**0.70** **0.82** **1.01**	**0.68** **0.82** 1.03	0.64 0.87 1.09	**0.71** **0.81** **0.99**	0.44 1.27 1.52	0.67 0.87 1.07	0.66 0.84 1.07	0.60 0.94 1.18	0.63 0.88 1.10	0.67 0.83 1.05	0.67 0.86 1.05	0.64 0.91 1.13	**0.68** **0.81** 1.04	0.63 0.90 1.11	**0.70** **0.82** **1.00**	0.61 0.88 1.12	**0.70** **0.82** **1.00**	0.62 0.90 1.11	CC MAE RMSE
Third experiment training/test split	**0.71** **0.69** **0.86**	0.67 0.72 0.92	**0.70** 0.72 0.89	**0.70** **0.70** **0.88**	0.66 0.77 0.96	**0.70** **0.70** **0.88**	0.66 0.72 0.93	0.63 0.76 0.95	**0.72** **0.69** **0.86**	0.48 1.24 1.48	0.69 0.74 0.95	0.64 0.77 1.01	0.59 0.83 1.11	**0.71** **0.70** **0.87**	**0.70** **0.69** 0.89	0.62 0.79 1.00	0.58 0.79 1.04	0.62 0.74 0.95	0.63 0.78 0.99	**0.70** **0.70** **0.88**	0.65 0.72 0.93	**0.70** **0.70** **0.88**	0.59 0.82 1.04	CC MAE RMSE
Overall performance	**0.80** **0.84** **1.05**	0.57 0.94 1.29	0.68 1.02 1.33	**0.75** 0.87 **1.09**	**0.76** 0.88 1.11	**0.75** **0.81** **1.00**	**0.76** **0.85** **0.95**	0.56 1.03 1.35	**0.79** **0.84** **1.02**	0.60 1.05 1.36	0.63 0.95 1.25	0.63 0.90 1.23	0.66 **0.86** 1.21	0.68 **0.86** 1.14	0.63 0.90 1.18	0.59 0.98 1.23	0.60 0.97 1.25	0.58 0.97 1.26	0.60 0.99 1.27	0.74 0.89 1.10	0.58 0.93 1.28	**0.75** 0.88 **1.09**	0.56 0.97 1.27	CC MAE RMSE

Fig. 11. Classifiers performance for arousal. GP – Gaussian Process; IR – Isotonic Regression; LMS – Least Mean Square; LR – Linear Regression; MP – Multilayer Perceptron; PR – Pace Regression; RBF – Radial Basis Function; SLR – Simple Linear Regression; SMO – SVM Regression; IBK – Instance-Based K-Nearest Neighbor; KS – K Star; LWL – Locally-weighted Learning; AR – Additive Regression; BAG – Bagging; ES – Ensemble Selection; RSS – Random SubSpace; RD – Regression By Discretization; CR – Conjuctive Rule; DT – Decision Table; M5R – M5 Rules; DS – Decision Stump; M5P – M5 Trees; REP – REP Tree

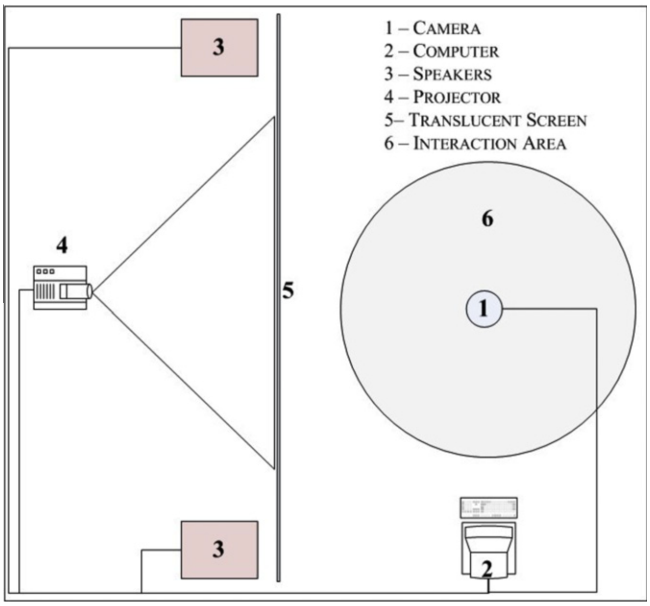

Fig. 12. Plan of the experimental setup.

questionnaires. Different listeners classified each emotional dimension of the segments with one value selected from the integer interval [0; 10]. Answers from listeners distant more than the mean ± 2 ∗ standard deviation (considered as outliers) were discarded.

In the first experiment we performed ad-hoc comparisons between a small group of classifiers [31], which allowed us to conclude that Support Vector Machine regression [55] obtained the best results. Because of this, results of this section were calculated with this type of classifier. We present an extended evaluation of various types of classifiers in the end of this section. The next subsections present details about each conducted experiment as well as the classification results with the application of 10-fold cross-validation: the correlation coefficients (CC), mean absolute errors (MAE), root mean square errors (RMSE), best features and their weights. The best features were selected with the help of the best first search method [55] and correspond to the features with the highest correlation with valence/arousal.

4.1. First experiment – preliminary evaluation of the classification module

The first experiment [31] aimed to identify the emotional relevance of 95 features (rhythmic, melodic, dynamic, textural, instrumental and harmonic). To accomplish this objective we analyzed the emotional answers of 53 listeners to 16 musical pieces. These pieces were of western tonal music (pop and r&b) and last between 20 and 60 s. Table 1 presents the results of this experiment.

4.2. Second experiment – extended evaluation of the classification module and analysis of audio features

The second experiment [32] extended the first by increasing the number of musical pieces (96), listeners (80) and features (322). Each listener tagged a subgroup of 16 pieces. Musical pieces were of western tonal music (film music) and last between 20 and 60 s. Table 2 presents the results.

Using the same data, we verified the importance of 18 audio features in the expression of emotions [33]. Spectral dissonance, spectral sharpness, spectral loudness [11] and spectral similarity were some of the features considered. Table 3 presents the best audio features of this experiment.

4.3. Third experiment – improvement of classification and transformation modules

The third experiment [35] was devoted to the verification of the effectiveness of the knowledge base in supporting the transformation algorithms and to make a subsequent update of the regression models. The test involved 132 pieces, 37 listeners and 337 features. Each listener tagged a subgroup of 22 pieces from a group of 132

Table 7
Seven 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

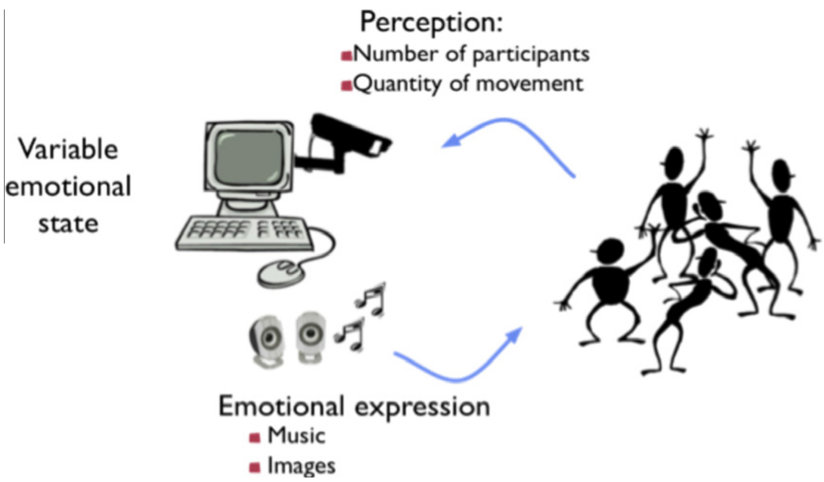

Fig. 13. Two main steps of the system: emotional perception and expression.

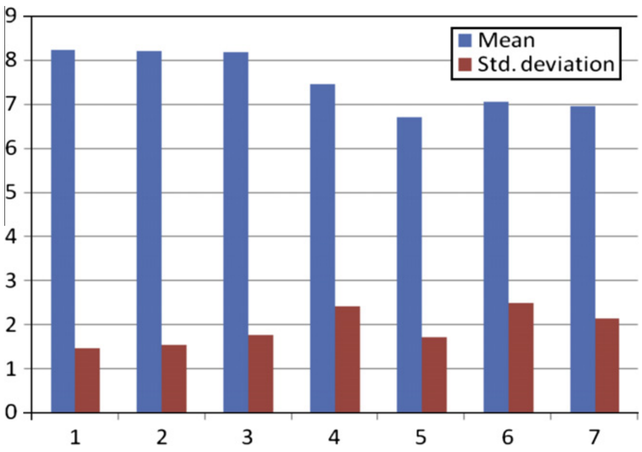

Fig. 14. Mean and standard deviations for the answers of the questionnaire.

musical pieces of western tonal music (pop/rock) that last between 10 and 15 s. Sixty-three of these pieces were used to test the effectiveness of transformation algorithms. The other 69 pieces were used to update regression models. Table 4 presents the results of this experiment.

4.4. Melodic analysis of the data obtained from all the experiments

Using the data from the experiments, we verified the importance of the melody in the expression of emotions. We manually extracted the melodic lines from the musical pieces used in the previous experiments. We guided this extraction by considering the loudness and pitch of the notes: notes with high loudness and pitch were considered as having a high probability of belonging to the melodic line. We used the listeners' answers obtained with the questionnaires of previous experiments and extracted from the melodic lines the same group of features that was used in the third experiment. Table 5 presents the results of this analysis.

We observe that there is variability on the best features for each of the experiments, particularly between those of the third experiment and those of the first two. The conditions that vary across the experiments are basically the style and the duration of the musical pieces. We believe that this variability may be explained by the differences in style. This will be investigated carefully in further experiences. We will come to this issue later on in Section 6.

4.5. Systematization of the data obtained from all the experiments

In the end of the four experiments, we collected the more discriminant (with a higher weight) features from each one (Tables 1–5) and obtained a group of 22/17 different features for valence/arousal. Then, we proceeded to a phase of feature selection by applying the best first search method and obtained a group of four features for valence and a group of four features for arousal. After this, we joined the musical and emotional data of the first, second and third experiments (Fig. 9) and proceeded to the application of 10-fold cross-validation with these groups of features (Table 6). We also calculated the percentage of correct predictions and obtained results of 79.0% for valence and 84.5% for arousal. We considered a correct prediction the one that falls in the interval of the mean value of the emotional answer, given by the listeners, plus or minus the standard deviation of this answer.

4.6. Evaluation of the performance of classifiers

With the systematization of the best features for each emotional dimension we were ready to evaluate the performance of various classifiers. We applied training/test split (66%/34%) and 10-fold cross-validation to evaluate the performance of several classifiers with their default parameters [55]. We used data of the first three experiments [31,32,35] and considered three metrics: correlation coefficient, mean absolute error, root mean square error. The classification of valence and arousal (Figs. 10 and 11) considered the best features of each experiment (Tables 1, 2 and 4). Concerning valence, support vector regression, least mean squares and regression by discretization obtained the best performances in the first experiment; linear regression, *M5R* and least mean squares obtained the best performances in the second experiment; linear regression, pace regression and support vector regression obtained the best performances in the third experiment; support vector regression, pace regression and linear regression obtained the best performances if we consider the mean of the results obtained in the three experiments.

Concerning arousal, Gaussian process, multilayer perception and support vector regression obtained the best performances in the first experiment; least mean squares, additive regression and bagging obtained the best performances in the second experiment; Gaussian process, support vector regression and linear regression obtained the best performances in the third experiment; support vector regression, Gaussian process and radial basis function obtained the best performances if we consider the mean of the results obtained in the three experiments. In brief, function-based models like support vector regression and Gaussian processes are the ones that perform better, whilst rule-based and tree-based models are the ones that perform worst. This may be explained by the robustness of the function-based models and lack of it on the other models.

4.7. Evaluation of the transformation module

Despite of the lower importance of the role of the transformation module when compared with the classification module, it was also subjected to experiments. This module also uses the regression models used by the classification module. The effectiveness of five of the six algorithms developed for this module was verified [34,35]. The transformation of tempo, note density, pitch register, spectral sharpness (Ambres), spectral sharpness (Zwickler), timbral width (spectral flatness) and loudness contributed to a direct influence on valence. The transformation of tempo, note density, spectral sharpness (Ambres), spectral sharpness (Zwickler) and spectral dissonance (Sethares) contributed to a direct influence on arousal. The transformation of pitch register and spectral similarity influenced arousal in an inverse way.

5. Application

Emotion-based interactive systems [10,15,36] have great application potential, namely in entertainment, engineering and healthcare. We developed an installation[5] to assess the interactive capabilities of EDME [49]. This installation is composed by a camera, a computer, speakers, a projector, a translucent screen and an interaction area (Fig. 12). The camera is placed on the ceiling of a room. The interaction area represented 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.

The computer has an emotional state (valence and arousal) that is defined according to the number of participants and quantity of movement. It has positive valence on the presence of people and negative valence when left alone. Moreover, people movement induces an increase in the arousal, whilst lack of activity induces a

[5] http://www.youtube.com/watch?v=Dodn_eOoBwo.

decrease. This computer selects music and images to express this state (Fig. 13). Music and images are expressed to the interaction area with the help of, respectively, the speakers and the translucent screen. At the interaction area, users can experience and influence the emotional behavior of the computer.

We made two informal experiments with the aim of obtaining feedback from the users. The first experiment had the objective of observing the interaction between the system and 30 participants. We obtained a positive verbal feedback: the system interacted in an expected emotional way. The second experiment was focused on obtaining the participants feedback, 23 in this case, via a questionnaire with seven questions (Table 7) about various components of the behavior of the system.

The answers obtained in the second experiment reveal some important features of the system: it correctly related arousal with the amount of movement and valence with the number of presences; music seems more important than images to express emotions; music was less successful than images in expressing the desired emotion; 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, as well as, by the presence of some questions with multiple components (e.g., first and second questions). Fig. 14 presents the mean and standard deviation for the seven questions of the questionnaire. They show us that the computer transmitted expected emotions and gains from using both music and images to express its emotional state.

6. Conclusion

EDME is a music production system that expresses a desired emotion. From its implementation resulted several advances to the state-of-the-art. It implements algorithms that control emotional content of music in different levels: segmentation, classification, selection and transformation. The knowledge base, one of the auxiliary structures, systematizes relations between emotions and musical features. It is also composed by an interface that allows different types of emotional representation. The flexibility of the architecture and the use of parameterizable structures widen the areas of application of EDME. The system was already applied in an affective installation. We also intend to demonstrate the usability of EDME in healthcare and soundtrack generation.

We used data obtained from web-based questionnaires to evaluate the performance of different classifiers; to update regression models being used; and to verify the effectiveness of transformation algorithms. We found a variability of the best features on each of the experiments, as well as on the section of melodic analysis. This may be explained by the fact that different styles of music are used on each of them. In the near future, we intend to validate/calibrate the system in controlled setups with different styles of music. One of the goals is to investigate whether it is possible to have one only regression model that covers all the styles. We are planning a set of experiments, focused on the emotions expressed by music, to collect data with questionnaires based on Self-Assessment Manikin [2]. At a later stage, we intend to assess the experienced emotions in listeners by collecting psychophysiological data and by recording facial expressions.

References

[1] D. Baum, EmoMusic – Classifying music according to emotion, in: Workshop on Data Analysis, 2006.

[2] M. Bradley, P. Lang, Measuring emotion: the self-assessment manikin and the semantic differential, Journal of Behavioral Therapy and Experimental Psychiatry (1994) 25:49–59.

[3] D. Cabrera, Psysound: a computer program for psychoacoustical analysis, in: Australian Acoustical Society Conference, vol. 24, 1999, 47–54.

[4] E. Cambouropoulos, Musical rhythm: a formal model for determining local boundaries, accents and metre in a melodic surface, Music, Gestalt, and Computing-Studies in Cognitive and Systematic Musicology (1997) 277–293.

[5] R. Cowie, Feeltrace: an instrument for recording perceived emotion in real time, in: Research Workshop on Speech and Emotion, 2000, pp. 19–24.

[6] J. Chung, G. Vercoe, The emotional remixer: personalized music arranging, in: Conference on Human Factors in Computing Systems, ACM Press,, New York, 2006, pp. 393–398.

[7] E. Daly, W. Lancee, J. Polivy, A conical model for the taxonomy of emotional experience, Journal of Personality and Social Psychology 45 (1983) 443–457.

[8] A. Damásio, S. Sutherland, Descartes' Error: Emotion, Reason and the Human Brain, Papermac London, 1996.

[9] D. Deutsch, The Psychology of Music, Academic Press, 1982.

[10] S. Dornbush, K. Fisher, K. McKay, A. Prikhodko, Z. Segall, Xpod: a human activity and emotion aware mobile music player. in: International Conference on Mobile Technology, Applications and Systems, 2005.

[11] T. Eerola, P. Toiviainen, Mir in Matlab: The Midi Toolbox, in: International Conference on Music Information Retrieval, 2004.

[12] P. Ekman, Basic Emotions, Handbook of Cognition and Emotion, 1999, pp. 45–60.

[13] A. Friberg, R. Bresin, J. Sundberg, Overview of the KTH rule system for musical performance, Advances in Cognitive Psychology 2 (2006) 145–161.

[14] N. Frijda, The Psychologists' Point of View. Handbook of Emotions, The Guilford Press, New York, 2000. pp. 59–74.

[15] H. Fujita, J. Hakura, M. Kurematu, Intelligent human interface based on mental cloning-based software, Knowledge-Based Systems 22 (3) (2009) 216–234.

[16] A. Gabrielsson, E. Lindström, The influence of musical structure on emotional expression, Music and Emotion: Theory and Research (2001) 223–248.

[17] Garritan Personal Orchestra. <http://www.garritan.com/GPO-features.html> (accessed 27.11.2009).

[18] P. Juslin, Communicating emotion in music performance: a review and a theoretical framework, Music and Emotion: Theory and Research (2001) 309–337.

[19] P. Juslin, P. Laukka, Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening, Journal of New Music Research 33 (3) (2004) 217–238.

[20] J. Larsen, G. Berntson, K. Poehlmann, T. Ito, J. Cacioppo, The psychophysiology of emotion, Handbook of Emotions (2008) 180–195.

[21] O. Lartillot, P. Toiviainen, MIR in Matlab (II): a toolbox for musical feature extraction from audio, in: International Conference on Music Information Retrieval, 2007, pp. 237–244.

[22] R. Lazarus, The Cognition-Emotion Debate: A Bit of History, Handbook of Cognition and Emotion, John Wiley & Sons Ltd., Sussex, 1999. 3-19.

[23] F. Lerdahl, R. Jackendoff, A Generative Theory of Tonal Music, MIT Press, 1996.

[24] E. Lindstrom, A Dynamic View of Melodic Organization and Performance, Ph.D. Thesis, Acta Universitatis Upsaliensis Uppsala, 2004.

[25] S. Livingstone, R. Muhlberger, A. Brown, A. Loch, Controlling musical emotionality: an affective computational architecture for influencing musical emotion, Digital Creativity (2007) 18.

[26] A. López, A. Oliveira, A. Cardoso, Real-time emotion-driven music engine, in: International Conference on Computational Creativity, 2010.

[27] C. McKay, I. Fujinaga, Jsymbolic: a feature extractor for Midi files, in: International Computer Music Conference, 2006.

[28] A. Mehrabian, Basic Dimensions for a General Psychological Theory, Cambridge OG&H Publishers, 1980.

[29] E. Narmour, The analysis and cognition of basic melodic structures: the implication-realization model, University of Chicago Press, 1990.

[30] A. Oliveira, A. Cardoso, Towards affective-psychophysiological foundations for music production, Affective Computing and Intelligent Interaction (2007) 511–522.

[31] A. Oliveira, A. Cardoso, Towards bidimensional classification of symbolic music by affective content, in: International Computer Music Conference, 2008.

[32] A. Oliveira, A. Cardoso, Modeling affective content of music: a knowledge base approach, in: Sound and Music Computing Conference, 2008.

[33] A. Oliveira, A. Cardoso, Emotionally-controlled music synthesis, Encontro de Engenharia de Áudio da AES Portugal (2008).

[34] A. Oliveira, A. Cardoso, Affective-driven music production: selection and transformation of music, in: International Conference on Digital Arts – ARTECH, 2008.

[35] A. Oliveira, A. Cardoso, Automatic manipulation of music to express desired emotions, in: Sound and Music Computing Conference, 2009.

[36] N. Oliver, F. Flores-Mangas, Mptrain: a mobile, music and physiology-based personal trainer, Conference on Human–Computer Interaction with Mobile Devices and Services, vol. 8, ACM Press, NY, 2006, pp. 21–28.

[37] A. Ortony, A. Collins, The Cognitive Structure of Emotions, Cambridge University Press, 1988.

[38] Project SAM Symphobia. <http://www.projectsam.com/Products/Symphobia/> (accessed 27.11.2009).

[39] J. Russell, Measures of emotion, Emotion: Theory, Research, and Experience 4 (1989) 83–111.

[40] K. Scherer, M. Zentner, Emotional effects of music: production rules, Music and Emotion: Theory and Research (2001) 361–392.

[41] K. Scherer, What are Emotions? And How Can They Be Measured?, Social Science Information 44 (4) (2005) 695–729

[42] E. Schubert, Measurement and Time Series Analysis of Emotion in Music, Ph.D. Thesis, University of New South Wales, 1999.

[43] J. Sloboda, Music structure and emotional response: some empirical findings, Psychology of Music 19 (1991) 110–120.
[44] A. Sorensen, A. Brown, Introducing JMusic, in: Australasian Computer Music Conference, 2000, pp. 68–76.
[45] Soundfont. <http://www.connect.creativelabs.com/developer/SoundFont/Forms/AllItems.aspx> (accessed 29.11.2009).
[46] T. Sugimoto, R. Legaspi, A. Ota, K. Moriyama, S. Kurihara, M. Numao, Modelling affective-based music compositional intelligence with the aid of ANS analyses, Knowledge-Based Systems 21 (3) (2008) 200–208.
[47] D. Temperley, The Cognition of Basic Musical Structures, MIT Press, 2004.
[48] K. Trohidis, G. Tsoumakas, G. Kalliris, I. Vlahavas, Multilabel classification of music into emotions, in: International Conference on Music Information Retrieval, 2008.
[49] F. Ventura, A. Oliveira, A. Cardoso, An emotion-driven interactive system, in: Portuguese Conference on Artificial Intelligence, 2009.
[50] K. Wassermann, K. Eng, P. Verschure, J. Manzolli, Live soundscape composition based on synthetic emotions, IEEE Multimedia 10 (2003) 82–90.
[51] S. Weisberg, Applied Linear Regression, Wiley-Blackwell, 2005.
[52] G. Widmer, W. Goebl, Computational models of expressive music performance: the state of the art, Journal of New Music Research 33 (3) (2004) 203–216.
[53] J. Wingstedt, M. Liljedahl, S. Lindberg, J. Berg, Remupp: An Interactive Tool for Investigating Musical Properties and Relations, New Interfaces For Musical Expression, University of British Columbia, 2005, pp. 232–235.
[54] R. Winter, Interactive Music: Compositional Techniques for Communicating Different Emotional Qualities, Master's Thesis, University of York, 2005.
[55] I. Witten, E. Frank, L. Trigg, M. Hall, G. Holmes, S. Cunningham, Weka: practical machine learning tools and techniques with java implementations, in: International Conference on Neural Information Processing, 1999, pp. 192–196.
[56] T. Wu, S. Jeng, Automatic emotion classification of musical segments, in: International Conference on Music Perception and Cognition, 2006.
[57] Y. Yang, Y. Lin, Y. Su, H. Chen, A regression approach to music emotion recognition, Audio, Speech, and Language Processing 16 (2008) 448–457.