Modeling Affective Content of Music: A Knowledge Base Approach

António Pedro Oliveira*, Amílcar Cardoso*
*Centre for Informatics and Systems of the University of Coimbra, Coimbra, Portugal

***Abstract*—The work described in this paper is part of a project that aims to implement and assess a computer system that can control the affective content of the music output, in such a way that it may express an intended emotion. In this system, music selection and transformation are done with the help of a knowledge base with weighted mappings between continuous affective dimensions (valence and arousal) and music features (e.g., rhythm and melody) grounded on results from works of Music Psychology.**

The system starts by making a segmentation of MIDI music to obtain pieces that may express only one kind of emotion. Then, feature extraction algorithms are applied to label these pieces with music metadata (e.g., rhythm and melody). The mappings of the knowledge base are used to label music with affective metadata. This paper focus on the refinement of the knowledge base (subsets of features and their weights) according to the prediction results of listeners' affective answers.

I. Introduction

Music has been widely accepted as one of the languages of emotional expression. The possibility to select music with an appropriate affective content can be helpful to adapt music to our affective interest. However, only recently scientists have tried to quantify and explain how music expresses certain emotions. As a result of this, mappings are being established between affective dimensions and music features [14][8].

Our work intends to design a system that may select music with appropriate affective content by taking into account a knowledge base with mappings of that kind. Most psychology researchers agree that affect has at least two distinct qualities [13][16][3]: valence (degree of satisfaction) and arousal (degree of activation), so we are considering these 2 dimensions in the classification. Automated classification using machine learning approaches has the advantage of allowing one to perform classifications in a faster and more reliable way than manual classifications. We intend to improve the knowledge base by selecting prominent features and by defining appropriate weights. This is done, respectively, by using feature selection and linear regression algorithms.

The automatic selection of music according to an affective description has a great application potential, namely in entertainment and healthcare. On the one hand, this system can be used in the selection of soundtracks for movies, arts, dance, theater, virtual environments, computer games and other entertainment activities. On the other hand, it can be used in music therapy to promote an intrinsic well-being. The next section makes a review of some of the most relevant contributions from Music Psychology and related works from Music Information Retrieval. Section III gives an overview of the system. Section IV presents the details of the experiment. Section V shows the experimental results. Section VI analyses the results, and finally, section VII makes some final remarks.

II. Related work

This work entails an interdisciplinary research involving Music Psychology and Music Information Retrieval. This section makes a review of some of the most relevant contributions for our work from these areas.

A. Music Psychology

Schubert [14] studied relations between emotions and musical features (melodic pitch, tempo, loudness, texture and timbral sharpness) using a 2 Dimensional Emotion Space. This study was focused on how to measure emotions expressed by music and what musical features have an effect on arousal and valence of emotions. Likewise, Korhonen [4] tried to model people perception of emotion in music. Models to estimate emotional appraisals to musical stimuli were reviewed [14][6] and system identification techniques were applied. Livingstone and Brown [8] provided a summary of relations between music features and emotions, in a 2 Dimensional Space, based on some research works of Music Psychology. Gabrielsson and Lindstrom [3] is one of these works, where relations between happiness and sadness, and musical features are established. Lindstrom [7] analysed the importance of some musical features (essentially melody, but also rhythm and harmony) in the expression of appropriate emotions.

B. Music Information Retrieval

Emotions detection in music can be seen as a classification problem, so the selection of the classifier model and the feature set are crucial to obtain good results. Van de Laar [17] compared 6 emotion detection methods in audio music based on acoustical feature analysis. Four central criteria were used in this comparison: precision, granularity, diversity and selection. Emotional expressions can be extracted from music audio [19]. The method designed by Wu and Jeng consisted in 3 steps: subject responses, data processing and segments extraction. From the results of this method, emotional content could be associated to musical fragments, according to some musical features like pitch, tempo and mode.

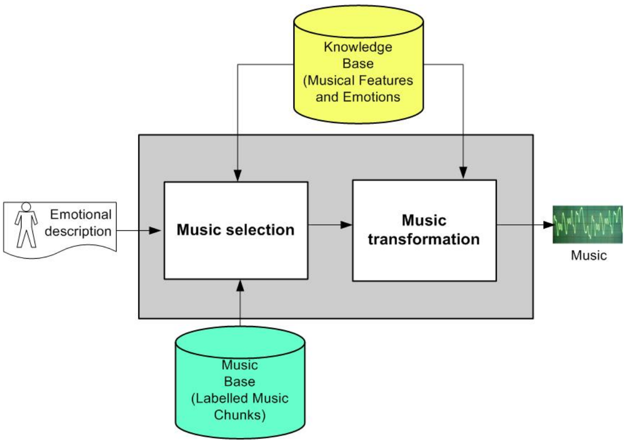

Fig. 1. System overview

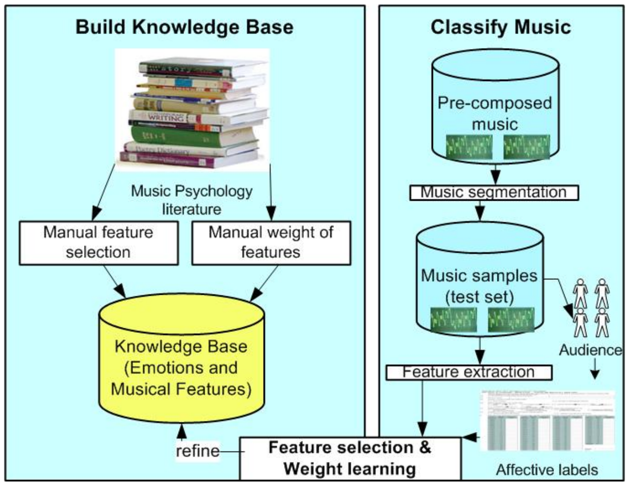

Fig. 2. Stages of the experiment

Muyuan and Naiyao [10] made an emotion recognition system to extract musical features from MIDI music. Support Vector Machines were used to classify music in 6 types of emotions (e.g., joyous and sober). Both statistical (e.g., pitch, interval and note density) and perceptual (e.g., tonality) features were extracted from the musical clips. There are also models to recommend MIDI music based on emotions [5]. The model of Kuo et al., based on association discovery from film music, proposes prominent musical features according to detected emotions. These features are compared with features extracted from a music database (chord, rhythm and tempo). Then, the result of these comparisons is used to rank music and a list of recommended music is given according to 15 groups of emotions.

III. Project description

The work described in this paper is part of a project that has the objective of implementing and assessing a computer system that can control the affective content of the music output, in such a way that it may express an intended emotion. The system uses a database of pre-composed music represented at a symbolic level. We intend to accomplish our objective in 2 stages. The first consists in the selection / classification of music by affective content and is the focus of this paper. The second stage will deal with the transformation of the selected music to approximate even further its affective content to an intended emotional description. These stages are done with the help of a knowledge base with weighted mappings between continuous affective dimensions (valence and arousal) and music features (e.g., rhythm and melody). Fig. 1 illustrates our system.

IV. Details of the experiment

The experiment here described follows a preliminary one [12], overcoming some of its limitations by using larger numbers of music files, listeners and music features. Fig. 2 presents an overview of different stages of our experiment. The process starts with the segmentation of MIDI music to obtain segments that may express only one kind of emotion (this method is described in detail in the following paragraph). Then, feature extraction algorithms of third party software [9][2][15] are applied to label these segments with music metadata (e.g., rhythm and melody). The mappings of the knowledge base are used to label music with affective metadata. For the experiments with listeners, we used a test set of 96 musical pieces. These pieces were of western tonal music (film music), last, approximately, from 20 seconds to 1 minute, and were used for both training and validating the classifier. 80 different listeners were asked to label online each affective dimension of the musical pieces with values selected from the integer interval between 0 and 10 [0;10][1]. The obtained affective labels were used to refine the sets of features and corresponding weights in the knowledge base. This was done separately for the valence and arousal.

A. Music segmentation

The expression of emotions in music varies as a function of time [4]. To facilitate classification, it is very important to obtain segments of music that may express only one kind of affective content. Our segmentation module uses the Local Boundary Detection Model (LBDM) [1][2] to obtain weights based on the strength of music variations (pitch, rhythm and silence). These weights establish plausible points of segmentation between segments with different musical features that may reflect different affective content. We define a threshold to reduce the search space among the LBDM weights. This threshold is equal to 1.30*mean(LBDM weights)+1.30*standard deviation(LBDM weights). We obtain music chunks of different length with a minimum of notes *MinN* and a maximum of notes *MaxN*. To segment, we start at the beginning of the MIDI file and look for a plausible point of segmentation that corresponds to the maximum weight between the beginning of MIDI file+*MinN* and the beginning of MIDI file+*MaxN*. This process is repeated starting from the last point of segmentation until we come to the end of the MIDI file.

[1] http://student.dei.uc.pt/%7Eapsimoes/PhD/Music/smc08/index.html

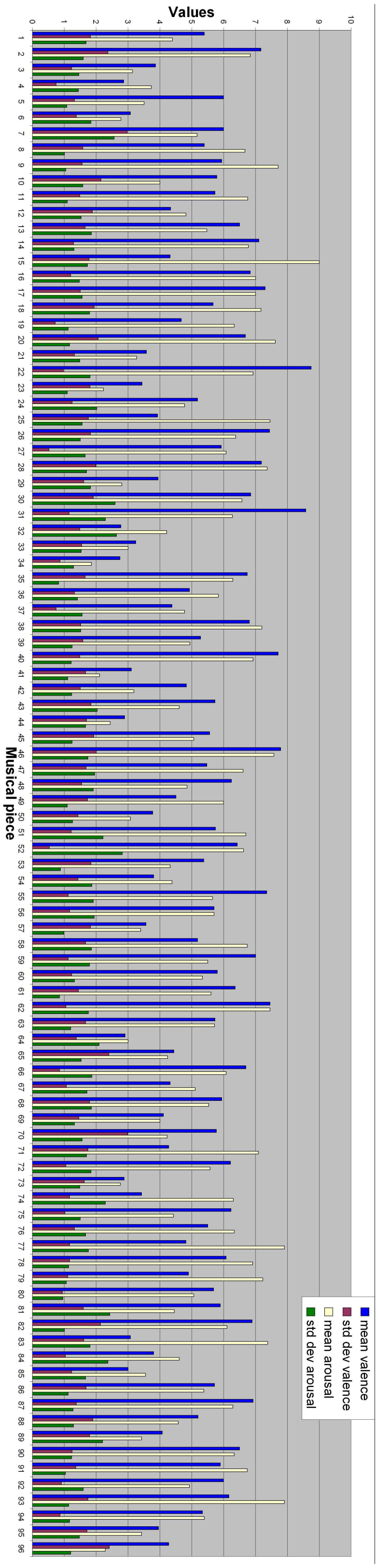

Fig. 3. Mean and standard deviations of the affective responses for valence and arousal

B. *Selection and weighting of features*

We examined 146 unidimensional features and 3 multidimensional ones that were categorized in 6 groups: instrumentation (20), texture (15), rhythm (39), dynamics (4), melody (68) and harmony (3). Special attention was devoted to new features and important ones from a preliminary experiment[12]: the importance (volume*time) of 13 MFCCs [17] of each sample used to synthesize musical instruments, the prevalence (by note or time) of specific groups and individual instruments, tempo, notes density, duration of notes, rhythmic variability, melodic complexity, number of repeated notes, prevalence of the most common melodic intervals, pitch classes and pitches, and mode (major or minor). Each feature was analysed for the affective dimensions with the help of the affective labels obtained for the test set and of information obtained from the literature [11]. This was done by applying the following feature selection algorithms: Genetic search, best-first and greedy stepwise [18].

With the subsets of features selected, some algorithms of linear regression were used to refine the weights of each feature. Linear regression, SMO regression and SVM regression [18] were tested. The results of the next section were obtained with SVM regression, because it was, generally, the approach that gave us the best results.

V. Results

Fig. 3 shows the mean and standard deviation for affective responses obtained in the online questionnaire[2]. Answers distant more than the mean $\pm$ 2*standard deviation were discarded.

The importance of individual features in each group of features was established (represented as positive or negative between parenthesis in tables I and II). All the features presented in tables I and II have a correlation coefficient higher than 15% with the affective labels.

A. *Valence*

Table I presents prediction results by groups of features for valence. From this, we can infer that rhythmic (e.g, tempo, average note duration, variability of note duration and time between onsets), harmonic (e.g., key mode and key), melodic (e.g., climax position and melodic complexity), texture (e.g., spectral texture MFCC 4 and 6, and number of unpitched instruments) and instrumentation features (e.g., string ensemble fraction) are relevant to the valence of music.

We started by applying feature selection algorithms [18] to reduce the number of features and to improve classification results. From this a group of 26 features resulted. The correlation and determination coefficients for training on the whole set were, respectively, 89.37% and 79.86%. 8-fold cross validation of classification resulted in correlation and determination coefficients of, respectively, 81.21% and 65.95%. After this we selected manually the best group of features to know the most important features in the stage of selection, but also for the stage of transformation. From this a group of 5 features

Features	Cor. Coef.	Det. Coef.
Note Prevalence Muted Guitar (+)	36.99%	13.68%
Electric Instrument Fraction (+)	33.72%	11.37%
Note Prevalence Steel Drums (+)	33.21%	11.02%
Time Prevalence Marimba (+)	31.41%	9.86%
Note Prevalence Fretless Bass (+)	31.02%	9.62%
Note Prevalence Timpani (-)	26.76%	7.16%
Electric Guitar Fraction (+)	23.4%	5.47%
String Ensemble Fraction (-)	21.5%	4.62%
Note Prevalence Pizzicato Strings (-)	21.08%	4.44%
Orchestral Strings Fraction (-)	20.7%	4.28%
Note Prevalence Orchestral Harp (-)	20.37%	4.14%
Saxophone Fraction (+)	19.75%	3.9%
Note Prevalence English Horn (-)	19.69%	3.87%
Note Prevalence French Horn (-)	19.56%	3.82%
Note Prevalence Tenor Sax (+)	19.18%	3.68%
Note Prevalence Synth Brass 1 (+)	19.12%	3.65%
Note Prevalence Pad 3 (polysynth) (+)	18.66%	3.48%
Note Prevalence Bassoon (-)	18.49%	3.41%
Time Prevale. Acoustic Grand Piano (-)	16.71%	2.79%
Acoustic Guitar Fraction (+)	16.46%	2.71%
Note Prevalence Ocarina (-)	16.18%	2.62%
Note Prevalence Banjo (-)	16.18%	2.62%
Note Prevalence Flute (-)	16.16%	2.61%
Woodwinds Fraction (-)	16.12%	2.60%
Note Prevalence Tuba (-)	15.88%	2.52%
Note Prevalence Xylophone (+)	15.0%	2.25%
Note Prevalence Accordion (+)	15.0%	2.25%
Spectral Texture MFCC 4 (+)	22.89%	5.23%
Spectral Texture MFCC 6 (+)	22.45%	5.04%
Spectral Texture MFCC 7 (+)	20.85%	4.35%
Number of Unpitched Instruments (+)	20.27%	4.11%
Spectral Texture MFCC 8 (+)	17.64%	3.11%
Spectral Texture MFCC 12 (-)	17.14%	2.94%
Number of Pitched Instruments (+)	16.39%	2.69%
Relative Note Density of Highest Line (-)	15.55%	2.42%
Initial Tempo (+)	62.95%	39.63%
Average Note Duration (-)	49.92%	24.92%
Average Time Between Attacks (-)	48.72%	23.73%
Strength Strong. Rhythmic Pulse (-)	42.72%	18.25%
Variability of Note Duration (-)	42.41%	17.98%
Note Density (+)	40.99%	16.8%
Strength Two Strong. Rhythmic Pulses (-)	37.66%	14.18%
Variability of Time Between Attacks (-)	36.57%	13.37%
Number of Relatively Strong Pulses (+)	30.24%	9.14%
Distinct Rhythm Count (+)	29.03%	8.43%
Rhythmic Variability (-)	28.06%	7.87%
Strength Sec. Strong. Rhythmic Pulse (-)	25.58%	6.54%
Strongest Rhythmic Pulse (+)	20.71%	4.29%
Average Meter Accent Synchrony (+)	19.88%	3.95%
Polyrhythms (-)	18.66%	3.48%
Staccato Incidence (+)	15.05%	2.26%
Climax Position (+)	32.7%	10.69%
Average Melodic Complexity (+)	24.15%	5.83%
Interval Strong. Pitch Classes (+)	20.84%	4.34%
Dominant Spread (+)	20.83%	4.34%
Consecutive Identical Pitches (+)	18.42%	3.39%
Key mode (-)	43.86%	19.23%
Key (-)	37.79%	14.28%
Strong Tonal Centres (-)	17.43%	3.04%

TABLE I
BEST FEATURES OF EACH GROUP - VALENCE

Features	Cor. Coef.	Det. Coef.
Electric Instrument Fraction (+)	28.48%	8.11%
String Ensemble Fraction (-)	27.79%	7.72%
Note Prevalence English Horn (-)	26.15%	6.84%
Number of Unpitched Instruments (+)	25.56%	6.53%
Note Prevalence Flute (-)	25.09%	6.29%
Brass Fraction (+)	25.0%	6.25%
Note Prevalence Orchestra Hit (+)	22.97%	5.28%
Electric Guitar Fraction (+)	21.5%	4.62%
Woodwinds Fraction (-)	21.08%	4.44%
Saxophone Fraction (+)	20.78%	4.32%
Percussion Prevalence (+)	20.75%	4.30%
Note Prevalence Tremolo Strings (+)	19.52%	3.81%
Note Prevalence Orchestral Harp (-)	18.96%	3.59%
Note Prevalence Electric Bass (finger) (+)	18.9%	3.57%
Time Prevalence Acoustic Guitar (nylon) (-)	17.79%	3.16%
Spectral Texture MFCC 2 (+)	28.16%	7.93%
Variab. Prevalence Unpitched Instruments (+)	25.86%	6.69%
Spectral Texture MFCC 4 (+)	24.82%	6.16%
Melodic Intervals in Lowest Line (-)	18.99%	3.61%
Relative Range of Loudest Voice (-)	17.84%	3.18%
Average Note Duration (-)	68.67%	47.15%
Note Density (+)	63.59%	40.44%
Variability of Note Duration (-)	57.4%	32.94%
Initial Tempo (+)	55.52%	30.82%
Average Time Between Attacks (-)	55.32%	30.6%
Variability of Time Between Attacks (-)	54.07%	29.23%
Average Duration Accent (-)	53.81%	28.95%
Strength Strongest Rhythmic Pulse (-)	47.58%	22.64%
Number of Relatively Strong Pulses (+)	43.86%	19.24%
Strength Two Strong. Rhythmic Pulses (-)	41.69%	17.38%
Polyrhythms (-)	38.33%	14.69%
Strongest Rhythmic Pulse (+)	35.51%	12.61%
Strength Second Strong. Rhythmic Pulse (-)	27.9%	7.78%
Onset Autocorrelation (-)	26.67%	7.11%
Syncopation (-)	25.36%	6.43%
Average Meter Accent Synchrony (-)	24.15%	5.83%
Number of Strong Pulses (+)	23.64%	5.59%
Rhythm Range (-)	23.46%	5.50%
Rhythmic Variability (-)	23.02%	5.30%
Staccato Incidence (+)	35.22%	12.40%
Average Range of Glissandos (-)	17.51%	3.07%
Climax Position (+)	45.39%	20.60%
Average Melodic Complexity (+)	38.4%	14.74%
Consecutive Identical Pitches (+)	37.06%	13.73%
Climax Strength (-)	33.12%	10.97%
Repeated Notes (+)	32.85%	10.79%
Most Common Pitch Class Prevalence (+)	31.59%	9.98%
Relative Strength of Top Pitch Classes (-)	30.69%	9.42%
Amount of Arpeggiation (+)	29.74%	8.84%
Same Direction Interval (+)	27.95%	7.81%
Repeated Pitch Density (+)	24.46%	5.98%
Most Common Pitch Prevalence (+)	24.39%	5.95%
Distance Common Melodic Intervals (+)	22.61%	5.11%
Overall Pitch Direction (+)	21.78%	4.74%
Most Common Melodic Interval Prevale. (+)	21.11%	4.46%
Melodic Octaves (+)	20.32%	4.13%
Melodic Thirds (-)	18.04%	3.25%
Interval Between Strongest Pitch Classes (+)	17.89%	3.2%
Duration of Melodic Arcs (+)	17.67%	3.12%
Key mode (-)	22.13%	4.90%

TABLE II
BEST FEATURES OF EACH GROUP - AROUSAL

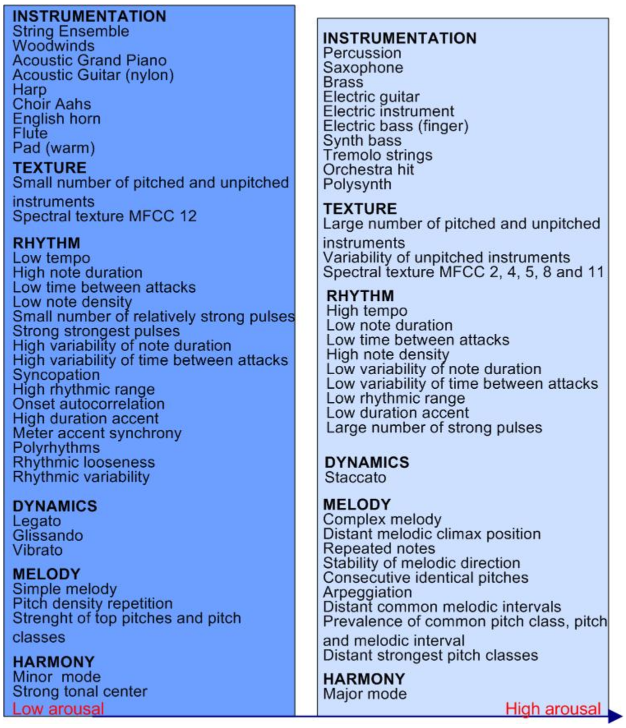

Fig. 4. Mappings for arousal

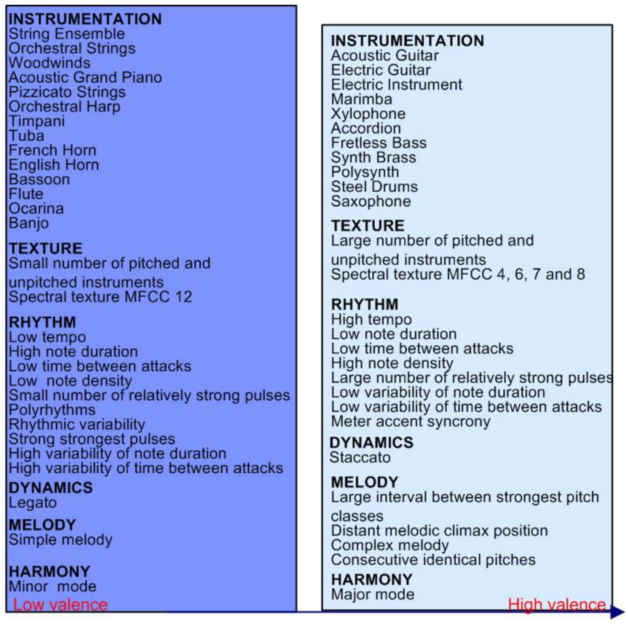

Fig. 5. Mappings for valence

resulted. The correlation and determination coefficients for training on the whole set were, respectively, 74.95% and 56.17%. 8-fold cross validation of classification resulted in correlation and determination coefficients of, respectively, 71.5% and 51.12%. Valence is calculated by the weighted sum of the best features: -0.41*average note duration + 0.17*dominant spread + 0.41*initial tempo - 0.18*key mode + 0.24*climax position.

Correlation and determination coefficients of 56.5% and 31.92% exists between the affective labels and results obtained using the weighted mappings of a preliminary experiment[1].

B. Arousal

Table II presents prediction results by groups of features for arousal. From this, we can infer that rhythmic (e.g., average note duration, note density, time between attacks and variability of note duration), dynamics (e.g., staccato incidence), texture (spectral texture MFCC 4 and strength of top pitch classes), melodic (e.g., climax position and repeated notes) and instrumentation features (e.g., number of unpitched instruments and brass fraction) are relevant to the arousal of music.

We started by applying feature selection algorithms [18] to reduce the number of features and to improve classification results. From this a group of 23 features resulted. The correlation and determination coefficients for training on the whole set were, respectively, 90.31% and 81.55%. 8-fold cross validation of classification resulted in correlation and determination coefficients of, respectively, 84.14% and 70.79%. After this we manually selected the best group of features to know the most important features in the stage of selection, but also for the stage of transformation. From this a group of 4 features resulted. The correlation and determination coefficients for training on the whole set were, respectively, 83.86% and 70.32%. 8-fold cross validation of classification resulted in correlation and determination coefficients of, respectively, 79.14% and 62.63%. Arousal is calculated by the weighted sum of the best features: -0.56*average note duration + 0.24*initial tempo + 0.11*climax position + 0.37*consecutive identical pitches + 0.58*note density.

Correlation and determination coefficients of 76.58% and 58.64% exists between the affective labels and results obtained using the weighted mappings of a preliminary experiment[1].

VI. Discussion

From this work much information was obtained. Fig. 4 and 5 illustrate how specific types of musical features can be changed to shift, respectively, the arousal and valence of music. For instance, a decrease in the duration of notes contribute to a decrease of both arousal and valence. Fig. 6 presents the mean of important features for each quadrant (e.g., high valence and low arousal). All this information is stored in the knowledge base and will be used in the next stage of our work that deals with the transformation of music affective content (Fig. 1).

With similar goals to [5] [10], we have developed a regression model (knowledge base) with relations between music features and emotions, Kuo et al. developed an affinity graph and Muyuan and Naiyao a SVM classifier. We used continuous dimensions (valence and arousal) instead of discrete emotions ([5] [10]). The results of our model ($\approx$ 90%) surpass the results of Kuo et al. ($\approx$ 80%) and Muyuan and Naiyao for valence ($\approx$ 70%) when using a higher number of features ($\approx$ 20).

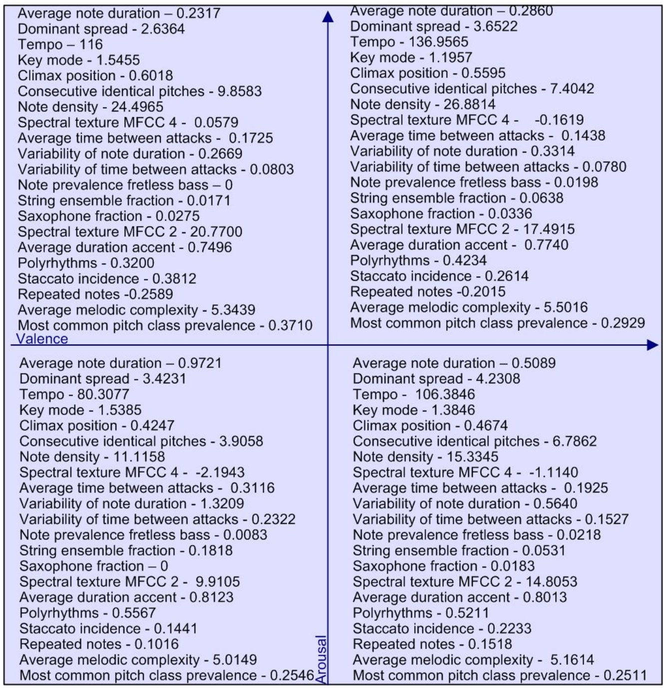

Fig. 6. Mean values of relevant features of musical samples for each affective quadrant

VII. Conclusion

We presented an extension of a previous work that undertook music emotion classification as a regression problem. SVM regression obtained the best results in the prediction and classification of the dimensions of valence and arousal. Validation results using the coefficient of determination confirmed that the prediction/classification of arousal (90.31%/81.55%) is easier than the prediction/classification of valence (89.37%/79.86%). Rhythmic (e.g., tempo, note density and average/variation of note duration), melodic (e.g., climax position and melodic complexity) and textural (e.g., spectral texture MFCCs) features proved to be very important to valence and arousal. Harmonic (e.g., key mode) and dynamics features (e.g., staccato incidence) were also important to predict, respectively, the valence and arousal. A correlation coefficient of 62.95% was obtained between valence and arousal.

With these satisfactory results, we feel ready to move to the second stage of our work, that consists in transformation of the affective content of selected music to approximate even further its affective content to an intended emotion. Both the selection and transformation will use the obtained information stored in the knowledge base (Fig. 4, 5, 6).

References

[1] Cambouropoulos, E. "The local boundary detection model (lbdm) and its application in the study of expressive timing.", *Int. Computer Music Conf.*, 2001.
[2] Eerola, T., Toiviainen, P. "Mir in matlab: The midi toolbox.", *International Conference on Music Information Retrieval*, 2004.
[3] Gabrielsson, A., Lindström, E. "The Influence Of Musical Structure On Emotional Expression." *Music and emotion: Theory and research*. Oxford University Press, 2001, 223–248.
[4] Korhonen, M. "Modeling continuous emotional appraisals of music using system identification." *Master's thesis*, University of Waterloo, 2004.
[5] Kuo, F., Chiang, M., Shan, M., Lee, S. "Emotion-based music recommendation by association discovery from film music." *ACM International Conference On Multimedia*, 2005, 507–510
[6] Li, T., Ogihara, M. "Detecting emotion in music." *International Conference on Music Information Retrieval*, 2003, 239–240
[7] Lindstrom, e. "A Dynamic View of Melodic Organization and Performance." *PhD thesis*, Acta Universitatis Upsaliensis Uppsala, 2004.
[8] Livingstone, S., Brown, A. "Dynamic response: real-time adaptation for music emotion." *Australasian Conference On Interactive Entertainment*, 2005, 105–111.
[9] McKay, C., Fujinaga, I. "jsymbolic: A feature extractor for midi files." *International Computer Music Conference*, 2006.
[10] Muyuan, W., Naiyao, Z., Hancheng, Z. "User-adaptive music emotion recognition." *International Conference on Signal Processing*, 2004.
[11] Oliveira, A., Cardoso, A. "Towards affective-psychophysiological foundations for music production." *Affective Computing and Intelligent Interaction*, 2007, 511–522.
[12] Oliveira, A., Cardoso, A. "Towards bi-dimensional classification of symbolic music by affective content." *International Computer Music Conference (to be published)*, 2008.
[13] Russell, J.A. "Measures of emotion." *Emotion: Theory, research, and experience*, 1989, 4, 83-111.
[14] Schubert, E. "Measurement and Time Series Analysis of Emotion in Music." *PhD thesis*, University of New South Wales, 1999.
[15] Sorensen, A. and Brown, A. "Introducing jMusic." *Australasian Computer Music Conference*, 2000, 68-76.
[16] Thimm K. and Fischer B. "Emotional responses to music: Influence of psycho-acoustical features." *National Conferences on Undergraduate Research*, 2003.
[17] van de Laar, B. "Emotion detection in music, a survey." *Twente Student Conference on IT*, 2006.
[18] Witten, I., Frank, E., Trigg, L., Hall, M., Holmes, G., Cunningham, S. "Weka: Practical machine learning tools and techniques with java implementations." *International Conference on Neural Information Processing*, 1999, 192–196.
[19] Wu, T., Jeng, S.: "Extraction of segments of significant emotional expressions in music." *Workshop on Computer Music and Audio Technology*, 2006.