Controling Music Affective Content: A Symbolic Approach

António Pedro Oliveira
Centre for Informatics and Systems, University of Coimbra, Portugal
apsimoes@student.dei.uc.pt - http://student.dei.uc.pt/~apsimoes

Amílcar Cardoso
Centre for Informatics and Systems, University of Coimbra, Portugal
amilcar@dei.uc.pt - http://eden.dei.uc.pt/~amilcar

In: C. Tsougras, R. Parncutt (Eds.)
Proceedings of the fourth Conference on Interdisciplinary Musicology (CIM08)
Thessaloniki, Greece, 2-6 July 2008, http://web.auth.gr/cim08/

Background in Music Computing. The automatic production of music according to an emotional description has a great application potential, namely in entertainment and healthcare. The control of the affective content may be accomplished by composing/arranging music, by selecting pre-composed music using adequate criteria, by transforming/adapting pre-composed music - currently this is only viable if working at a symbolic representation level - or by combinations of the above.

Background in Music Psychology. Music has been widely accepted as one of the languages of emotional expression. However, only recently scientists have tried to quantify and explain how music influences our emotions. Relations between emotions and musical features have been studied from different perspectives: physiological, psychological, sociological, historical, mathematical, etc. Physiological reactions to music stimuli, the role of music features in emotional expression, the meaning of emotions in music, and models of emotional expression in music are some relevant topics.

Aims. We intend to implement and assess a computer system that can control the affective content of pre-composed music represented at a symbolic level, in such a way that produced music is adapted to the intended emotional description.

Main contribution. We are implementing and assessing a system that automatically produces music with an appropriate affective content. The implementation consists in the development of algorithms for the segmentation, analysis, selection, transformation, sequencing, remixing and synthesis of pre-composed symbolic music. This is done with the help of a knowledge base with weighted mappings between continuous affective dimensions (valence and arousal) and musical features (rhythm, melody, etc.) grounded on results from works of Music Psychology. The assessment of the system consists in playing produced music and identifying emotions of the listener.

Our system starts with the reception of an emotional description specified by the listener. The mappings more suitable to the emotional description are selected from the knowledge base. These mappings are used in the stages of selection and transformation of music to foster a good control of the affective content in music. Music is selected from a music base (MIDI files), according to similarity metrics between music features (timbre, rhythm, etc.). Afterward, selected music can be subject to transformation, sequencing, remixing and synthesis algorithms. Produced music is played in experimental contexts (e.g., clinical or entertainment), where data is collected about the emotions of the listeners by using questionnaires and/or psychophysiological data. To calibrate the system identified emotions are compared with the intended emotions. These comparisons are used to refine the mappings in the knowledge base. New mappings can be added; old mappings can be adapted or discarded. Both Case-Based and Rule-Based techniques are known to be adequate for this kind of operation.

Implications. From a general standpoint, this work has particular significance to the fields of Music Psychology and Music Computing. Our computational systematization of mappings between emotions and music features can be helpful for scientific research in Music Psychology. Moreover, both structural (e.g., harmonic mode and overall pitch) and performing features (e.g., melody accent and beat accent) can be controlled (selected and transformed) in the production of music with an appropriate affective content. Thus, our system can be used by musicians as an affective music production tool or as an autonomous affective DJ-like application.

With this system an appropriate expression of an emotional experience can be tailored by using music. Possible applications include the production of soundtracks for arts, movies, dance, theater, virtual environments, computer games and other entertainment activities. Another area of application is the use of produced music as therapeutic mean to promote an intrinsic wellbeing.

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 (Livingstone and Brown, 2005; Schubert, 1999). Our work intends to design a system that may control music affective content by taking into account a knowledge base with weighted mappings of that kind. Feature selection and linear regression algorithms are being used to, respectively, select prominent features and define appropriate weights for these mappings, which are essential for the algorithms of selection and manipulation of our system.

The automatic production 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 production of soundtracks for movies, arts, dance, theater, virtual environments, computer games and other entertainment activities. On the other hand, it can be used to produce music characterized by an affective content of tenderness, calm, love, joy, peace, kindness and goodness that could promote an intrinsic wellbeing. The next section makes a review of some of the most relevant contributions from Music Computing and Music Psychology. Next, we give an overview of each part of our work. Later, we present the assessment methodology, and finally, we make some final remarks.

Background

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

Music Computing

The automatic production of music according to an emotional description has a great application potential, namely in entertainment and healthcare. The control of the affective content may be accomplished in 4 different approaches.

The first approach consists in composing/arranging music. Winter (2005) built a real-time application to control structural factors of a composition. Models of musical communication of emotions were reviewed to get an insight of what musical features are relevant to express emotions. Pre-composed musical scores were manipulated through the application of rules with control values for different musical features: mode, instrumentation, rhythm and harmony. Rutherford (2003) evaluated musical pieces produced by a rule-based real-time music generator and identified a non-uniform relation between scariness and the tension perceived in the music. MAgentA (Casella and Paiva, 2001) is an agent that automatically produces real-time background music for a virtual environment. This environment has an emotional state, which is given to the agent. The emotional state is used to select an algorithm from a database of affective composition algorithms. These algorithms try to match the emotional state to musical features (harmony, melody, tempo, etc.).

The second approach consists in selecting pre-composed music using adequate criteria. Music classification is helpful for the selection of music with similar features. Muyuan et al. (2004) 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 (Kuo et al., 2005). The model of Kuo et al., based on association discovery from film music, proposed prominent musical features according to a queried emotion description. These features were compared with features extracted from a music database (chord, rhythm and tempo). Then, music was ranked and a music list was recommended according to 15 groups of emotions.

The third approach consists in transforming/adapting pre-composed music - currently this is only viable if working at a symbolic representation level. REMUPP (Wingstedt et al., 2005) is a system that allows real-time manipulation of music features like tonality, mode, tempo and instrumentation. Pre-composed music is given to a music player and specific music features are used to control the sequencer (e.g., tempo), to employ filters and effects (e.g., rhythmic complexity) and to control synthesizers (e.g., instrumentation). Livingstone and Brown (2005) implemented a rule-based system to affect perceived emotions of music, by modifying the musical structure. This system is grounded on a list of performative and structural features, and their emotional effect.

The fourth and last approach consists in combining mentioned approaches. Chung and Vercoe (2006) developed a system to generate real-time music based on intended listener's affective expressions with the aim to correlate musical parameters with changes in affective state. Music files were generated in real-time by music composition/production, segmentation and re-assembly of music. There was also a model to select music segments and remix them to induce an appropriate emotion. The analysis of listener's affective state was based on physiological data (skin arousal), physical data (foot tapping) and a questionnaire.

Music Psychology

Music is an artistic mean for communicating/expressing our feelings. Different scientific perspectives have been used to quantify and explain how music influences our emotions: physiological, psychological, sociological, historical, mathematical, etc. Scherer and Zentner (2001) established parameters of influence (structural, performance, listener and contextual) for the experienced emotion. Physiological reactions like shivers, laughter, tears and lump in the throat are the result to musical phenomena like melodic appogiaturas, unexpected harmonies, crescendos syncopation and enharmonic changes (Sloboda, 1991). Meyer (1956) analyzed structural characteristics of music and its relation with emotional meaning in music. Continuity, completeness, uniformity, expectation and variation of musical features were analysed from an emotional perspective. The way the vertical (pitch – harmony, instrumentation and texture) and horizontal (temporal – rhythm, melody and dynamics) features are organized in music have an influence on the affective content of music (Koelsch and Siebel, 2005). Some works have tried to measure emotions expressed by music and to identify the effect of musical features on emotions (Gabrielsson and Lindstrom, 2001; Korhonen, 2004; Lindstrom, 2004; Schubert, 1999). From this, relations can be established between emotions and musical features (Livingstone and Brown, 2005).

Work Description

We are implementing and assessing a system that can control the affective content of pre-composed music represented at a symbolic level, in such a way that produced music is adapted to the intended emotional description. MIDI files are being used instead of audio files because they allow us to analyse and transform easier high-level musical features. Pre-composed music is obtained from websites and subject to algorithms of segmentation that intend to allow a better control of affective content by reducing the size of musical pieces. Then, feature extraction algorithms are applied to label these pieces with music metadata (e.g., rhythm and melody) which are then stored in a music base.

Music selection and manipulation 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. Music selection intends to obtain musical pieces with an affective content similar to the intended emotion. Then, these pieces can be transformed to approximate even further the affective content. Sequencing and remixing algorithms intend to put the obtained pieces in an ordered and fluent way. Music synthesis consists in selecting the samples that best fit the sequenced music. Next paragraphs are dedicated to the presentation of these processes in more detail with the aid of Figure 1.

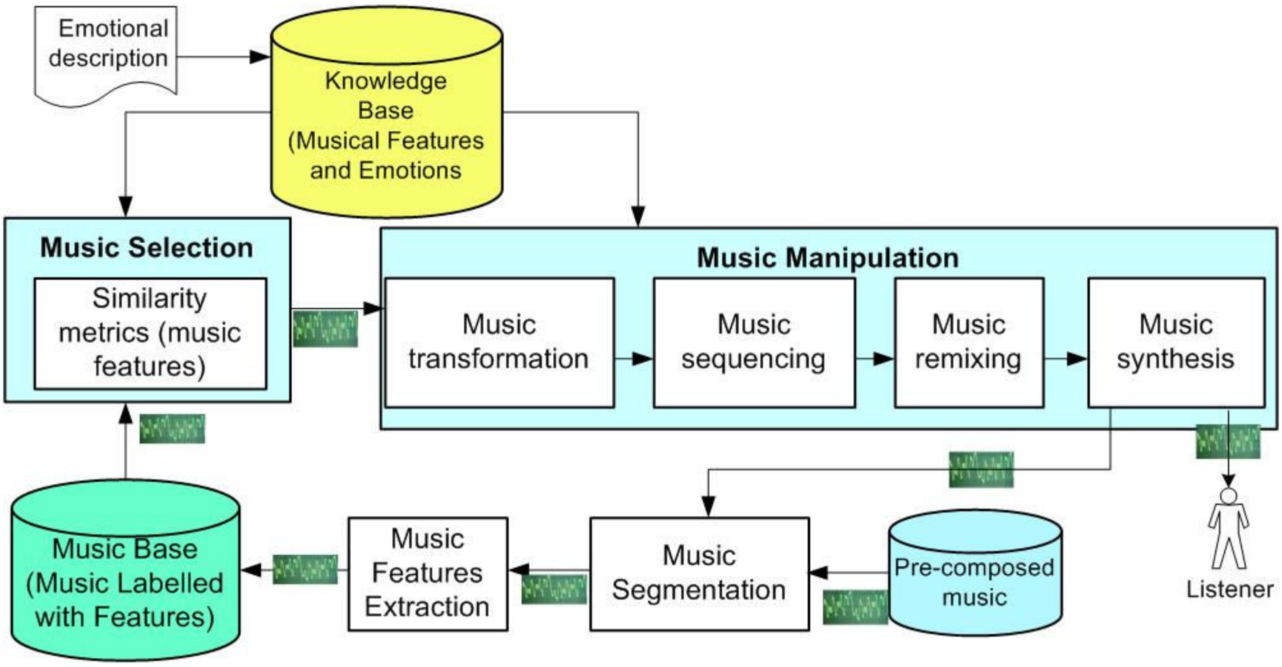

Figure 1. Loop of control of music affective content with the help of a knowledge base.

Music segmentation

The expression of emotions in music varies as a function of time (Korhonen, 2004). So, to facilitate classification it is very important to obtain segments of music that express only one kind of affective content. Our segmentation module uses the Local Boundary Detection Model (Eerola and Toiviainen, 2004) to obtain weights for plausible points of segmentation. Pre-composed music is segmented into music chunks of different length defined by the number of notes. We define a threshold to reduce the search space. This threshold is equal to 1.30*mean(weights) + 1.30*standard deviation(weights). Chunks have 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.

Music features extraction

The extraction of features is done with the help of third party software: JSymbolic (McKay and Fujinaga, 2006), MIDI toolbox (Eerola and Toiviainen, 2004) and JMusic (Sorensen and Brown, 2000). 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).

Each feature was analysed for the affective dimensions with the help of affective labels of 100 listeners for 100 music samples. Special attention was devoted to features with a high degree of correlation with the labels: the importance (volume*time) of 13 MFCCs 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).

Knowledge base

The knowledge base models the affective content of music with weighted mappings between continuous affective dimensions (valence and arousal) and music features (e.g., rhythm and melody) grounded on a literature review of music research works (Oliveira and Cardoso, 2007).

The knowledge base is being improved by selecting prominent features and by defining appropriate weights for each affective dimension. This is being done, respectively, by using feature selection and linear regression algorithms. Features are selected with the help of the following algorithms: Genetic search, best-first and greedy stepwise. With the features selected, weights are refined with the help of the following algorithms: linear regression, SMO regression and SVM regression.

Music selection

To make easy the process of music selection we are automatically classifying music by affective content using the knowledge base. The emotional output of each music is calculated through a weighted sum of the features, with the help of a vector of weights for each affective dimension (Figure 2).

$$Valence = \sum_{i=0}^{n} valenceWeight_i * feature_i$$

$$Arousal = \sum_{i=0}^{n} arousalWeight_i * feature_i$$

Figure 2 Equations for music affective content.

Music transformation

Selected music can be subject to transformations to approximate even further its affective content to an intended emotion. Our system has performative rules and structural rules to change affective content of music. We are designing algorithms to manipulate instrumentation (e.g., instruments fraction and timbre), musical texture (e.g., number of pitched and unpitched instruments) rhythm (e.g., note density, note duration, tempo and time between attacks), melody (e.g., climax position, melodic complexity, average pitch and consecutive identical pitches), dynamics (e.g., loudness, staccato incidence and range of glissandos) and harmony (e.g., mode and scale).

Musical features have a different impact on the affective content of music. Tempo and note density seems to have a higher importance for valence than the direction of melodic motion and timbre, so higher weights are given to tempo and note density. To control timbre we are analyzing features of audio samples (e.g., MFCCs and register). Clustering techniques are then used to group samples with similar timbre features. The knowledge base is used to help in selecting the sample that best match the intended emotion and the musical features of the MIDI track. Note density can be increased/decreased by adding/deleting tracks. For instance, drum track can be deleted to decrease note density and a specific track can be duplicated to increase note density. These transformations have a side-effect on musical texture features. Valence can be increased by increasing the number of consecutive identical pitches. Changing music mode from minor mode to major mode also has a positive effect on valence. Other harmonic transformations can also be done depending on the consonance of music intervals.

Music sequencing

Sequencing of music includes the ordering of tracks by an intended affective content, but also by musical features, namely tempo. To make easier the process of sequencing we are applying clustering techniques to group musical pieces with similar musical features (e.g., rhythm and harmony). Crossfading techniques are used to synchronize the pitch, tempo, and phase of two sequenced tracks.

Music remixing

Remixing algorithms are a special case of sequencing algorithms when two remixed tracks have some points of overlapping. Clustering techniques can also be used but they have to take into account that the MIDI tracks are being put in parallel and not in a sequential way (like in music sequencing).

Music synthesis

Soundfonts are being used to synthesize musical instruments. The synthesis process is controlled by analyzing features of audio samples (e.g., MFCCs and register). The way music synthesis is done is related to the way we intend to transform timbre, so that it may approximate the affective content of produced music to the intended emotion.

Control features of the knowledge base

According to the experiments that we have done there are specific features with a higher weight. To the arousal of music, we inferred 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 valence of music, we inferred 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.

Assessment methodology

Figure 3 presents the assessment methodology used in this system. Intended emotional descriptions are used to select appropriate mappings from the knowledge base. According to these mappings, music is played in experimental contexts (e.g., clinical or entertainment), where data is collected about the emotions of the listeners by using questionnaires and/or psychophysiological data. To calibrate the system comparisons are made between identified and intended emotions. These comparisons are used to refine the mappings in the knowledge base. New mappings can be added; old mappings can be adapted or discarded. Both Case-Based and Rule-Based representations are known to be adequate for this kind of operation.

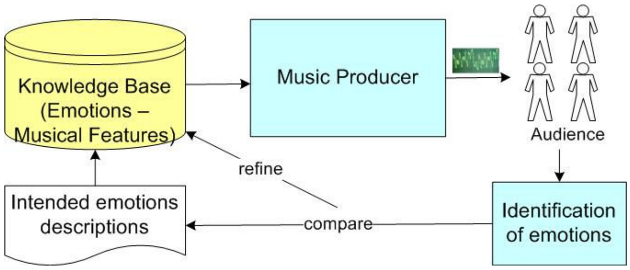

Figure 3 Loop of control of the assessment methodology.

Conclusion

We are building a computational model of music production that may express intended emotions. To accomplish this we are implementing and assessing a system that uses a computational systematization of mappings between emotions and musical features to control the affective content of music in the segmentation, analysis, selection and manipulation of music.

With this system calibrated an appropriate expression of an emotion can be tailored by using music. Possible applications include the production of soundtracks for arts, movies, dance, theater, virtual environments, computer games and other entertainment activities. Another area of application is the use of produced music as a therapeutic mean to promote an intrinsic wellbeing.

Acknowledgments. Thanks go to all people who answered to the online questionnaire, namely to the members of Center for Informatics and Systems of the University of Coimbra.

References

Casella, P. and Paiva, A. (2001). Magenta: An architecture for real time automatic composition of background music.

International Workshop on Intelligent Virtual Agents. London, UK, Springer-Verlag, 224–232.

Chung, J. and Vercoe, G. (2006). The affective remixer: Personalized music arranging. *Conference on Human Factors in Computing Systems*. ACM Press New York, NY, USA, 393–398.

Eerola, T. and Toiviainen, P. (2004). Mir in matlab: The midi toolbox. *International Symposium on Music Information Retrieval*.

Gabrielsson, A. and Lindstrom, E. (2001). The Influence Of Musical Structure On Emotional Expression. *Music and emotion: Theory and research*. Oxford University Press, 223–248.

Koelsch, S. and Siebel, W. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences* 9, 578–584.

Korhonen, M. (2004). *Modeling continuous emotional appraisals of music using system identification.* Master's thesis, University of Waterloo.

Kuo, F., Chiang, M., Shan, M. and Lee, S. (2005). Emotion-based music recommendation by association discovery from film music. *ACM International Conference On Multimedia*, 507–510.

Lindstrom, e. (2004). *A Dynamic View of Melodic Organization and Performance*. PhD thesis, Acta Universitatis Upsaliensis Uppsala.

Livingstone, S. and Brown, A. (2005). Dynamic response: real-time adaptation for music emotion. *Australasian Conference On Interactive Entertainment*, 105–111.

McKay, C. and Fujinaga, I. (2006). Jsymbolic: A feature extractor for midi files. *International Computer Music Conference*.

Meyer, L. (1956). *Emotion and Meaning in Music*. University of Chicago Press.

Muyuan, W., Naiyao, Z. and Hancheng, Z. (2004). User-adaptive music emotion recognition. *International Conference on Signal Processing*.

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

Rutherford, J. and Wiggins, G. (2003). An experiment in the automatic creation of music which has specific emotional content. *MOSART: Workshop On Current Research Directions In Computer Music*, 35–40.

Scherer, K. and Zentner, M. (2001) Emotional effects of music: Production rules. *Music and emotion: Theory and research*, 361–392.

Schubert, E. (1999). *Measurement and Time Series Analysis of Emotion in Music*. PhD thesis, University of New South Wales.

Sloboda, J. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music* 19(2), 110.

Sorensen, A. and Brown, A. (2000). Introducing jMusic *Australasian Computer Music Conference,* 68-76.

Wingstedt, J., Liljedahl, M., Lindberg, S. and Berg, J. (2005). Remupp: An interactive tool for investigating musical properties and relations. *New Interfaces For Musical Expression*. University of British Columbia, Vancouver, Canada, 232–235.

Winter, R. (2005). *Interactive music: Compositional techniques for communicating different emotional qualities*. Master's thesis, University of York.