Affective-Driven Music Production: Selection and Transformation of Music

António Pedro Oliveira and Amílcar Cardoso

Centre for Informatics and Systems, 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 helped by a knowledge base with weighted mappings between music features (e.g., note duration and tempo) and continuous affective dimensions (valence and arousal) grounded on mappings from music research literature and experimentally-derived weighted mappings. This paper focus on the evaluation of these mappings in the selection and transformation of music for the expression of an intended emotion.**

***Index Terms* — Artificial intelligence, Computer music.**

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 [8, 15]. Our work intends to design a system that may control music affective content by taking into account a knowledge base with mappings derived experimentally [13] and grounded on music research literature [12].

The automatic production of music according to a desired emotion 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 entertainment activities. On the other hand, it can be used to produce music characterized by an affective content of tenderness, calm, love, joy and peace that could promote an intrinsic wellbeing. The next section makes a review of some of the most relevant contributions from computational and scientific areas of music research. Next, we give an overview of the system. Later, we present the evaluation of the system in the selection and transformation of music for the expression of an intended emotion, and finally, we make some final remarks.

II. RELATED WORK

This work entails an interdisciplinary research involving a computational (music computing) and scientific (music psychology) perspective. This section makes a review of some of the most relevant contributions for our work from these areas.

The automatic production of music according to an intended emotion 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. This can be done by controlling structural factors of a composition [19] or by selecting an algorithm, from a database of affective composition algorithms that best match musical features to a desired emotion [1]. The second approach consists in selecting pre-composed music using adequate criteria. This can be done by making a recomendation model (e.g., graph, SVM) using features extracted from music (e.g., statistical and perceptual) [6, 11, 20]. The third approach consists in transforming/adapting pre-composed music - currently this is only viable if working at a symbolic representation level. This can be done by manipulating music features of pre-composed music using appropriate rules of control [8, 18]. The fourth approach consists in combining some of the mentioned approaches [2].

Different scientific perspectives have been used to quantify and explain how music expresses emotions: physiological, psychological, sociological, historical, mathematical, etc. These perspectives support and show prominent areas of exploration for the computational control of affective content in music. Physiological reactions like shivers, laughter, tears and lump in the throat result from musical phenomena like melodic appoggiaturas, unexpected harmonies, crescendos and syncopation [16]. Structural characteristics of music are related to the emotional meaning in music [10]. Continuity, completeness, uniformity, expectation and variation of musical features were analyzed 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 has an influence on the affective content of music [4]. Grounded on these findings, some works have measured emotions expressed by music [5, 7, 8, 15].

III. System 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 emotion. MIDI files are being used instead of audio files because they allow us to analyze and transform easily high-level music features. Pre-composed music is obtained from websites and can be 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 employed to obtain music metadata (e.g., rhythm and melody) that is used to label these pieces which are then stored in a music base.

Music selection and transformation are done with the help of a knowledge base with weighted mappings between music features (e.g., rhythm and melody) and continuous affective dimensions (valence and arousal) grounded on mappings from music research literature [12] and experimentally-derived weighted mappings [13]. 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 desired affective content. Music synthesis is sample-based, and our approach includes the selection of samples according to the music transformation objectives. Next paragraphs are dedicated to the presentation of these processes in more detail with the aid of Fig. 1.

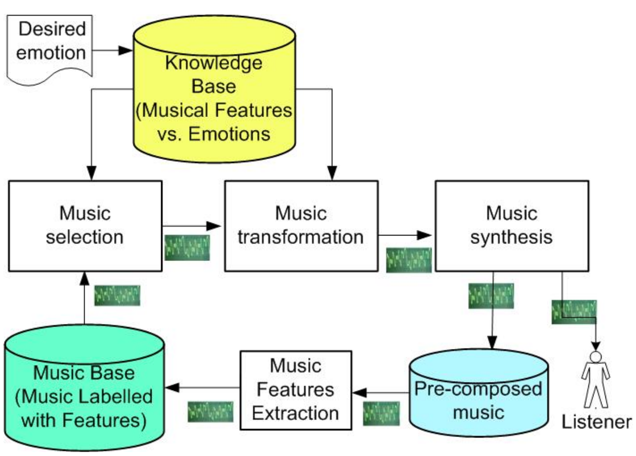

Fig. 1. System architecture for affective-driven music production

A. Music features extraction

The extraction of features was constrained by the available features in third party software: Jsymbolic [9] MIDI toolbox [3] and Jmusic [17]. We examined the importance of 146 unidimensional features and 3 multidimensional ones for the emotional expression [14]. These features belong to 6 groups: instrumentation, texture, rhythm, dynamics, melody and harmony.

B. Knowledge Base

The knowledge base models the affective content of music with weighted mappings between music features (e.g., note duration and tempo) and continuous affective dimensions (valence and arousal). Weights have a value x in $\Re : x \in$ [0; 1]. We evaluated features in the discrimination of the affective content and as a result of this 10 music features are being used to control the selection and transformation of the affective content:

- Average note duration – represents the average duration (in seconds) of the notes of all MIDI tracks. This feature has a weight of -0.99 for valence and -0.81 for arousal.
- Importance of bass register – represents the fraction of notes with pitch between 0 and 54. This feature has a weight of -0.34 for valence.
- Importance of high register – represents the fraction of notes with pitch between 73 and 127. This feature has a weight of -0.09 for valence.
- Initial tempo – represents the tempo in beats per minute. This feature has a weight of 0.29 for valence and 0.23 for arousal.
- Note density – represents the average number of notes per second. This feature has a weight of 0.45 for valence and 0.65 for arousal.
- Percussion prevalence – represents the total number of notes corresponding to percussion instruments divided by total number notes. This feature has a weight of 1 for arousal.
- Repeated notes – represents the fraction of notes that are repeated in each of the MIDI tracks. This feature has a weight of 0.33 for arousal.
- Variation of dynamics – represents the standard deviation of loudness levels of all notes. This feature has a weight of 0.15 for valence.
- Key mode – represents the mode of the music (major or minor). This feature has a weight of -0.74 for valence.
- Note prevalence of instruments – represents the total number of notes corresponding to each General MIDI (GM) Instrument divided by total number notes. Tables I and II shows the weights of each GM instrument, respectively, for valence and arousal.

TABLE I

WEIGHTS OF GM INSTRUMENTS FOR VALENCE

GM Instrument	Weights for valence							
1-8	0	0	1	-1	-1	1	1	1
9-16	-1	1	-1	1	1	1	1	0
17-24	-1	1	-1	0	0	1	-1	0
25-32	0	0	0	1	1	-1	-1	0
33-40	0	0	0	0	0	0	0	0
41-48	-1	-1	-1	1	1	1	-1	1
49-56	-1	-1	-1	-1	-1	-1	-1	1
57-64	1	1	1	1	0	1	1	1
65-72	0	0	0	0	-1	-1	-1	-1
73-80	-1	-1	1	1	1	1	1	-1
81-88	0	0	0	0	-1	0	0	0
89-96	0	0	1	0	0	0	0	0
97-104	0	0	0	0	0	0	0	0
105-112	0	0	0	0	1	0	0	0
112-120	1	1	1	1	0	0	0	0
121-128	0	0	0	0	0	0	0	0

TABLE II

WEIGHTS OF GM INSTRUMENTS FOR AROUSAL

GM Instrument	Weights for arousal							
1-8	0	-1	-1	0	0	-1	0	0
9-16	-1	-1	-1	0	-1	-1	0	0
17-24	0	0	0	0	0	1	0	1
25-32	0	0	0	0	0	1	1	1
33-40	1	1	1	1	1	1	1	1
41-48	-1	-1	-1	1	1	-1	-1	1
49-56	-1	-1	-1	-1	0	0	0	1
57-64	1	1	1	0	0	1	1	1
65-72	0	0	0	0	0	-1	0	-1
73-80	-1	-1	0	0	0	0	-1	0
81-88	1	1	0	0	1	0	1	1
89-96	-1	-1	1	0	0	0	0	0
97-104	0	0	-1	0	0	0	0	0
105-112	0	0	0	0	-1	1	1	0
113-120	0	-1	0	-1	0	0	0	0
121-128	0	0	0	0	0	0	0	0

C. Music selection

Music is automatically classified by affective content using the knowledge base, and selection is done according to this affective content. The affective content of each music is calculated through a weighted sum of the features, with the help of a vector of weights for each affective dimension:

$$Valence = \sum_{i=0}^{n} valenceWeight_i * feature_i \quad (1)$$

$$Arousal = \sum_{i=0}^{n} arousalWeight_i * feature_i \quad (2)$$

D. Music transformation

Selected music can be subject to transformations to approximate even further its affective content to an intended emotion. Our system has the possibility to change the affective content by changing specific music features. We have implemented algorithms to transform the mentioned list of 10 features, except for repeated notes, variation of dynamics and key mode. Note density can be increased/decreased by adding/deleting tracks (e.g., a drum track can be deleted to decrease note density, and specific tracks can be duplicated with different instruments to increase note density). To increase the importance of the high/bass register, music is transposed up/down by a specific number of octaves. To control timbre (and its prevalence on music), we are acting in the music synthesis process, as described next.

E. Music synthesis

Soundfonts are being used to synthesize musical instruments. As we intend to transform timbre, so that it may approximate the affective content of produced music to the intended emotion, the synthesis process is controlled by the weights of each GM instrument (tables I and II), which were calculated according to the importance of features of the audio samples (e.g., MFCCs, ADSR envelope and register) in the emotional expression [13, 14]. Clustering techniques are used to group samples with similar timbre features. The knowledge base helps in selecting the sample that best matches the intended emotion and musical features of the MIDI track.

IV. SYSTEM EVALUATION

We are evaluating the regressive model obtained in previous work [13, 14]. For these experiments we used a set of 13 pieces of pop/rock music to evaluate the algorithms of selection and 11 pieces of classical music to evaluate the algorithms of transformation. Musical pieces last, approximately, from 20 seconds to 6 minutes and can be listened in the following website[1]. Two listeners were asked to label each affective dimension of the musical piece with values x selected from $N : x \in$ [0; 10]. The obtained affective labels were used to evaluate both types of algorithms, separately, for the valence and arousal.

A. Valence

For selection we obtained a correlation and determination coefficients of, respectively, 87.44% and 76.46%. For transformation we obtained a correlation and determination coefficients of, respectively, 67.75% and

[1] http://papersao.googlepages.com/

45.90%. Table III shows the valence obtained with the listeners and system for the music used in selection and transformation.

B. Arousal

For selection we obtained a correlation and determination coefficients of, respectively, 84.84% and 71.98%. For transformation we obtained a correlation and determination coefficients of, respectively, 69.40% and 48.16%. Table III shows the arousal obtained with the listeners and system for the music used in selection and transformation.

TABLE III
VALENCE/AROUSAL OF THE LISTENERS AND SYSTEM

Valence				Arousal			
Selection		Transform.		Selection		Transform.	
List.	Syst.	List.	Syst.	List.	Syst.	List.	Syst.
9	7.4	2	5.8	7	6.3	3	3.7
6	6.5	5	7.9	6	6.5	4	5.2
8	7.7	5	6.4	7	7.3	4	4.3
9	7.5	4	7.2	8	7.4	3	3.5
7	7.2	9	8.9	7	6.1	5	3.9
8	7.9	4	4.7	7	6.8	3	2.9
8	7.3	3	1.7	7	7.1	3	0.1
9	7.5	7	5.4	7	7.2	4	5.2
9	7.7	6	6.7	6	5.9	5	4.6
7	6.8	3	5.1	6	5.7	3	2.0
6	6.7	8	9.5	5	4.6	5	6.9
5	6.0	-	-	5	5.7	-	-
6	6.9	-	-	8	7.2	-	-

V. CONCLUSION

We are building computational model of music production that may express intended emotions. To accomplish this we are implementing and evaluating a system that uses a computational systematization of mappings between musical features and emotions to control the affective content of music in the selection and transformation of music. Correlations coefficients were calculated for valence and arousal: 87.4% and 84.8% in the selection, and 67.8% and 69.4% in the transformation. These results, if confirmed by statistically relevant tests, show a good degree of control over the affective content.

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 and other entertainment activities. Another area of application comprises a therapeutic use of produced music to promote intrinsic wellbeing.

REFERENCES

[1] 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.

[2] 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.

[3] Eerola, T. and Toiviainen, P. (2004). “Mir in matlab: The midi toolbox.” *Int. Conf. on Music Information Retrieval*.

[4] Koelsch, S. and Siebel, W. (2005). “Towards a neural basis of music perception.” *Trends in Cog. Sci.* 9, 578-584.

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

[6] 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.

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

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

[9] McKay, C. and Fujinaga, I. (2006). “Jsymbolic: A feature extractor for midi files.” *Int. Computer Music Conference*.

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

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

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

[13] Oliveira, A., Cardoso, A. (2008). ”Towards bi-dimensional classification of symbolic music by affective content” *Int. Computer Music Conference*.

[14] Oliveira, A., Cardoso, A. (2008). ”Modeling Affective Content of Music: A Knowledge Base Approach” *Sound and Music Computing Conference*.

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

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

[17] Sorensen, A. and Brown, A. (2000). “Introducing jMusic.” *Austr. Comp. Music Conf.*, 68-76.

[18] 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.

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

[20] Yang, Y.H., Lin, Y.C., Su, Y.F. and Chen, HH (2008). “A Regression Approach to Music Emotion Recognition.” *IEEE Transactions on Audio, Speech, and Language Processing*,448-457