EMOTION-FLOW GUIDED MUSIC ACCOMPANIMENT GENERATION

Yi-Chan Wu and Homer H. Chen

National Taiwan University, Taiwan
Emails: {b99901175, homer}@ntu.edu.tw

ABSTRACT

The emotion of a music piece varies as it unrolls in time. We develop a system that takes a melody and an expected emotion flow as input and automatically generates an accompaniment. The accompaniment is composed of chord progression and accompaniment pattern. The former is generated from melody and valence data through dynamic programming, and the latter from arousal data. A mathematical model is developed to describe the relation between valence and chord progression. The performance of the system is evaluated subjectively. The cross-correlation coefficient between the expected arousals and the perceived ones is 0.84, and the cross-correlation coefficient between the expected valences and the perceived ones is 0.52. Both coefficients exceed 0.90 for musician subjects.

***Index Terms*—**Accompaniment, emotion flow, music emotion, melody harmonization

1. INTRODUCTION

Music plays an important role in human history; there is no known culture that lacks music [1]. Most music pieces consist of melody and accompaniment [2]. However, different accompaniments for a melody can evoke quite different emotional responses from listeners [3]. Since the affection evoked by music in a listener builds up and falls repeatedly as the music unrolls in time, we aim to develop an automatic accompaniment generation system that is able to create such dynamic music emotion perception.

The input to our system includes a melody and a pair of valence and arousal curves specified as a function of time by the user. Music emotion is effected through the manipulation of music features. Mode and chord features are related to valence perception [4]–[6], therefore, we embed them in chord progression and use chord progression to control valence. In contrast, we control arousal by manipulating rhythm, volume, and pitch range [5], [6] through the accompaniment pattern.

The contributions of the paper are as follows:

- The proposed system automatically generates an affective accompaniment that conforms to the valence and arousal sequences and the melody input by the user (Section 4).
- We develop a model to describe the relation between valence and chord progression. The optimization problem of valence-based chord progression generation is solved by dynamic programming (Section 5).

2. RELATED WORK

Several automatic accompaniment generation systems have been developed in the last few years. The accompaniment systems described in [7], [8] harmonize the melody by a hidden Markov model that represents the melody as an observation sequence and the chord as a hidden state. Similarly, in [9], a Markov chain with neo-Riemannian transformations is used to construct chord progression. However, these systems do not incorporate affection into accompaniment.

There is limited research on the generation of affective accompaniments for melodies. In MySong [10], [11], different chord transition matrices are used to generate the accompaniment according to the "happy factor" provided by the user. The system described in [12] generates the affective accompaniment based on the concept that the same chord followed by different chords evokes different emotions. These systems only generate accompaniments with one single emotion, not a flow of emotions.

Affective music synthesis is a relevant research topic. Most affective music synthesis systems [4], [5] generate the melody after the chord progression is determined. Therefore, the music is much easier to harmonize. Our system works in a reverse fashion. It generates the chord progression for a given a melody. Since the melody serves as a constraint, the generation of chord progression becomes an optimization problem, for which both melody and emotion have to be considered.

3. EMOTION AND MUSIC FEATURES

In this section, we first describe the emotion model adopted in this work and the two important dimensions, valence and arousal. Then we describe the music features relevant to emotion perception.

3.1. Emotion model

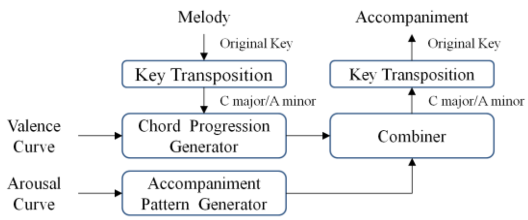

Fig. 1. System block diagram.

Various emotion models have been proposed in the past decades. A typical model is the one that classifies emotion into distinct categories. However, the number of categories is too small in comparison with the full spectrum of human emotion [13]. Therefore, an alternative approach that represents emotion by continuous values in a multi-dimensional space has been considered.

In music emotion, Thayer's model [14] is most commonly adopted. It has two attribute axes: valence (how positive/negative) and arousal (how exciting/calming). With this model, our system generates an accompaniment according to the valence and arousal curves specified by the user.

3.2. Music features relevant to valence and arousal

A general consensus is that mode, chord, and melodic direction are music features that are relevant to the valence perception, while tempo, rhythm, volume, and pitch range are relevant to the arousal perception [4]–[6].

However, not all of such music features are applicable to our system. For example, the tempo and the melodic direction are already decided by the melody and cannot be altered. Therefore, we use the other features to generate affective accompaniments. Since the mode and chord features are embedded in chord progression, we may simply focus on how to express valence with chord progression. The model we develop to relate valence to chord progression is described in Section 5.4.

In our system, arousal is expressed by manipulating rhythm, volume, and pitch range. These music features are controlled through the accompaniment pattern. Different chords played with the same accompaniment pattern sound similar in rhythmic structure. The design of accompaniment patterns is described in Section 6.

4. SYSTEM OVERVIEW

Fig. 1 shows the block diagram of the proposed system. The input to the system is a melody, a valence curve, and an arousal curve.

First, the system uses MIDI toolboxes [15], [16] to read the MIDI file of the melody and analyze the melodic features. Then, it transposes the melody into C major (A minor) if the key is in major (minor). In our system, relative major and minor keys are jointly considered, so only the

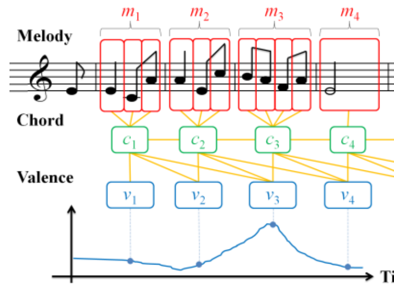

Fig. 2. Given a valence value v_i and its corresponding melody bar m_i, the problem of chord progression generation is to determine an appropriate chord c_i that, together with c_{i-1}, c_{i-2}, …, and c_{i-N}, where N is an integer (2 in this figure), would evoke v_i from the listener.

chords frequently used in both C major and A minor, namely, C, Dm, Em, F, G, and Am [17], are concerned.

Second, for each bar of the melody, the system finds an appropriate chord from the chord database to meet the following three requirements: 1) The transition of chords has to be smooth and natural, 2) Each chord should harmonize with the corresponding bar of the melody, and 3) The chord progression should conform to the given valence curve. The three requirements are considered in Section 5.

The system generates the accompaniment pattern according to the given arousal curve. For each bar of the melody, the system selects a pattern from the pattern database (Section 6) and assigns it to the chord progression for accompaniment generation. Finally, the accompaniment is transposed to the original key and combined with the melody to form a MIDI file.

5. GENERATING CHORD PROGRESSION

The generation of chord progression is formulated as an optimization problem using three cost functions. For practical purposes, a suboptimal solution is developed to trade optimality for computational efficiency.

5.1 Optimization formulation

The formulation of chord progression generation problem is illustrated in Fig. 2, where m_i denotes the ith bar of the melody, c_i the corresponding ith chord, and v_i the corresponding ith sample of the input valence curve. Also let $\mathbf{m}$, $\mathbf{c}$, and $\mathbf{v}$ denote their vector forms, each of size n. Each requirement described in Section 4 is expressed as a cost function. Denote the three cost functions by f_1, f_2, and f_3. Then the chord progression generation is formulated as

$$\mathbf{c}^* = \arg\min_{\mathbf{c}} \lambda_1 f_1(\mathbf{c}) + \lambda_2 f_2(\mathbf{m}, \mathbf{c}) + \lambda_3 f_3(\mathbf{v}, \mathbf{c}) \qquad (1)$$

where $\mathbf{c}^*$ denotes the optimal chord progression and λ_1, λ_2, and λ_3 are positive constants.

5.2 Cost function $f_1(\mathbf{c})$

In Western music, some chords are more likely to follow other chords [17]. Such chord transitions make the music sound smooth and natural. Therefore, the first cost function is defined by

$$f_1(\mathbf{c}) = -\log_2 p(\mathbf{c}) \approx -\sum_{i=1}^{n} \log_2 p(c_i \mid c_{i-1}) \qquad (2)$$

where $p(\cdot)$ denotes probability and n, as defined before, denotes the number of melody bars. The approximation assumes the transition of chords is a first-order Markov process [7], [8]. In our system, we calculate the transition matrix based on the Theorytab database [18], which contains approximately 6,000 songs with various genres. The cost is inversely proportional to the probability.

5.3 Cost function f_2(m, c)

If a melody bar predominantly contains the pitches that make up a chord, then the chord harmonizes well with the bar. We define the second cost function as follows:

$$f_2(\mathbf{m}, \mathbf{c}) = \sum_{i=1}^{n} f_2(m_i, c_i) = -\sum_{i=1}^{n} \log_2[\mathbf{P}(m_i) \cdot \mathbf{T}(c_i)] \qquad (3)$$

where $\mathbf{P}(\cdot)$ denotes the pitch class profile of a melody bar and $\mathbf{T}(\cdot)$ denotes the template of a chord. The former represents the proportion of the 12 pitches in the chromatic scale [19], whereas the latter represents the binary information of the pitches in a chord [20]. For example, in Fig. 2, we have $\mathbf{P}(m_1)$ = [1/4, 0, 0, 0, 1/2, 0, 0, 0, 0, 1/4, 0, 0]. If the first chord is an A minor chord, we have $\mathbf{T}(c_1 = \text{Am})$ = [1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0] since it consists of three notes, A, C, and E. The inner product of the two vectors is 1, which means the chord is perfectly harmonized with the melody.

5.4 Cost function f_3(v, c)

We develop a model to relate chord progression to valence based on the observation that the same chord followed by different chords evokes different emotions [12]. Let $\Delta v_{c,i}$ denote the valence value of two successive chords (c_{i-1}, c_i), and $\mathbf{\Delta v_c}$ its vector form. We determine $\Delta v_{c,i}$ by

$$\Delta v_{c,i} = g(c_{i-1}, c_i) = \begin{cases} 1, & \text{if } c_i \text{ is a major chord} \\ -1, & \text{if } c_i \text{ is a minor chord} \end{cases} \qquad (4)$$

where $g(\cdot)$ denotes a function. In general, the function can be subjectively determined by a ranking-based method [12]. We take a simplified approach based on empirical data.

Besides, we assume that the latest N+1 chords, or the corresponding N chord pairs, determine the perception of the current valence value (Fig. 2) and that a more recent chord has a higher impact on the current valence. Thus, we have

$$\mathbf{v_c} = \mathbf{\Delta v_c} * \mathbf{w}, \qquad (5)$$

where $\mathbf{w} = \frac{2}{N(N+1)}\left[\, N, N-1, ..., 2, 1 \,\right]$ is a filter. We pick N = 7 for the system.

Fig. 3. Accompaniment patterns shown with C major chords.

Then we calculate the square error between $\mathbf{v}$ and $\mathbf{v_c}$ as the third cost function,

$$f_3(\mathbf{v}, \mathbf{c}) = \| \mathbf{v} - \mathbf{v_c} \|^2 = \sum_{i=1}^{n} (v_i - v_{c,i})^2. \qquad (6)$$

5.5 Dynamic programming

If we ignore the emotion part by setting $\lambda_3 = 0$, (1) becomes an ordinary harmonization problem, which can be easily solved by dynamic programming. This case can be realized by erasing the lines between **v** and **c** in Fig. 2. The dynamic programming process starts from the first chord, stores the least cost of all possible current ending chords, and moves on to the next chord. It has been extensively used for melody harmonization, for which the melody is considered an observation sequence and the chord a hidden state in a hidden Markov model [7], [8], [10].

In general, (1) is more complex than the melody harmonization problem. It is practically infeasible to find the optimal chord progression by the same method since all 6^7 (approximately 280,000) results for the last seven chords have to be stored in each iteration. The complexity becomes greater if a larger chord database is used.

To solve the problem, we only store the results for the last two chords. Since the weights of these chords are much larger, this suboptimal solution leads to reasonably good results.

6. GENERATING ACCOMPANIMENT PATTERN

The accompaniment pattern is generated from the given arousal curve based on three music features: rhythm, volume, and pitch range. Fig. 3 shows the eight accompaniment patterns generated by the system. The note density increases rapidly in the first four patterns, whereas strong bass notes appear in the last four patterns due to contrast enhancement. We can see that the arousal increases from left to right and from top to bottom.

Let a_i denote the ith sample of the input arousal curve. In our system, we quantize a_i using a uniform eight-level quantizer and choose the accompaniment pattern for the corresponding bar m_i.

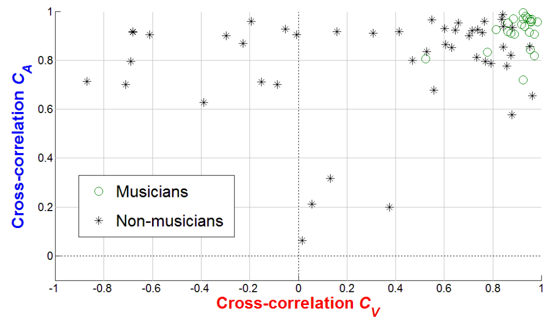

Fig. 4. Distribution of the cross-correlation coefficients for musicians (green circles) and non-musicians (black asterisks).

7. SYSTEM EVALUATION

A subjective test is set up to evaluate the performance of our system. The details are described in this section.

7.1 Setup

A total of 12 subjects are recruited, four musicians (one female and three males, mean age 23.0) and eight non-musicians (four females and four males, mean age 22.6). Three melodies are used in the subjective test. To reduce burden, each melody is shorter than 60 seconds.

In the test, each subject specifies the expected emotion flow by drawing a pair of valence and arousal curves for each melody. This step is repeated twice, and the results are averaged. Each curve is drawn as a function of time and within the range [–1, 1]. Altogether, the system generates six accompaniments. Then, as the system plays back the accompanied melody, the subject is asked to annotate its perceived emotion as a function of time and within [–1, 1]. Repeatedly listening is allowed.

Two restrictions are imposed on the user-specified emotion curves. First, no more than four peaks and valleys (excluding the two endpoints) are allowed for each curve. This is reasonable, considering the duration of each melody. Second, each curve should have a certain extent of variation, because the cross-correlation, which is used to measure the similarity between two emotion curves, becomes unstable if the curves are flat. The variation requirement can be expressed as

$$\frac{1}{T}\int_0^T \left(v(t)-\overline{v}\right)^2 dt > 0.1 \text{ and } \frac{1}{T}\int_0^T \left(a(t)-\overline{a}\right)^2 dt > 0.1 \quad (7)$$

where $\overline{v}$ and $\overline{a}$ denote the means of the specified curves $v(t)$ and $a(t)$, and T is the duration of the melody. A system warning is given if the requirement is not met. In this case, the subject has to redraw the emotion curve.

7.2 Evaluation method

In practice, we cannot expect the subject to generate exactly the same emotion curve for the same music when listening to it the second time. Therefore, the Pearson correlation, which is invariant to separate changes in location and scale between two curves, is a more appropriate similarity metric than the mean square error for this work. In addition, we allow a two-second lag in calculating the cross-correlation to account for the lag in response [21], [22].

Table 1. Statistics of cross-correlation coefficients between arousal curves

C_A	All	Musicians
Mean	0.8373	0.9181
Std. dev.	0.1837	0.0675

Table 2. Statistics of cross-correlation coefficients between valence curves

C_V	All	Musicians
Mean	0.5199	0.9017
Std. dev.	0.5224	0.0951

7.3 Evaluation results

Fig. 4 shows the results of the subjective test, one score per subject per accompanied melody. Totally, there are 72 scores in the score chart. The horizontal axis represents the cross-correlation coefficient C_V between a user-specified valence curve and its corresponding perceived valence curve. The vertical axis C_A is defined in the same way for arousal.

We can see a number of scores with negative C_V's in the score chart. This is consistent with the fact that valence perception is more subjective than arousal perception [23]. Nevertheless, about two thirds of the scores have $C_V > 0.60$, and all except five scores have $C_A > 0.60$, indicating that the system is capable of generating reasonably good accompaniments for the users.

The results for musicians are remarkable. There are 21 out of 24 scores with $C_V > 0.80$ and $C_A > 0.80$, and the means of both C_V and C_A exceed 0.90. This means our valence-based chord progression model works effectively well for musicians. The superior scores obtained for musicians can be attributed to the music training they have received. In general, musicians are more sensitive to subtle changes in music [24]. More statistics are shown in Tables 1 and 2.

8. CONCLUSION

In this paper, we have described an automatic system that generates an affective accompaniment according to the emotion flow specified by the user. The development of the system is based on a valence model for chord progression and a set of accompaniment patterns to express arousal. We have performed a subjective test to evaluate the system, and the results show that the accompaniments generated by the system do express the emotions specified by the users.

9. REFERENCES

[1] D. Huron, “Is music an evolutionary adaptation?” *Annals of the New York Academy of Science*, vol. 930, pp. 43–61, 2001.

[2] M. Lewis, J. M. Haviland-Jones, and L. F. Barrett, *Handbook of Emotions*, 3rd ed. NewYork: Guilford Press, 2010, ch. 7.

[3] R. Parncutt, *Harmony: A Psychoacoustical Approach*. NewYork: Springer-Verlag, 1989, ch. 4–5.

[4] I. Wallis, T. Ingalls, E. Campana, and J. Goodman, “A rule-based generative music system controlled by desired valence and arousal,” in *Proc. SMC*, 2011.

[5] F. Morreale, R. Masu, and A. De Angeli, “Robin: An algorithmic composer for interactive scenarios,” in *Proc. SMC*, 2013.

[6] P. Gomez and B. Danuser, “Relationships between musical structure and psychophysiological measures of emotion.” *Emotion*, vol. 7, no. 2, pp. 377–387, 2007.

[7] M. Allan and C. K. I. Williams, “Harmonising chorales by probabilistic inference.” *Advances in Neural Information Processing Systems*, vol. 17, pp. 25–32, 2004.

[8] S. A. Raczynski, S. Fukayama, and E. Vincent, “Melody harmonization with interpolated probabilistic models,” *Journal of New Music Research*, vol. 42, no. 3, pp. 223–235, 2013.

[9] C.-H. Chuan and E. Chew, “A hybrid system for automatic generation of style-specific accompaniment,” in *Proc. Int. Joint Workshop on Computational Creativity*, pp. 57–64, 2007.

[10] I. Simon, D. Morris, and S. Basu, “MySong: Automatic accompaniment generation for vocal melodies,” in *Proc. CHI*, pp. 725–734, 2008.

[11] D. Morris, I. Simon, and S. Basu, “Exposing parameters of a trained dynamic model for interactive music creation,” in *Proc. AAAI*, Chicago, pp. 1–8, 2008.

[12] P.-C. Chen, K.-S. Lin, and H. H. Chen, “Emotional accompaniment generation system based on harmonic progression.” *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1469–1479, 2013.

[13] Y.-H. Yang and H. H. Chen. “Machine recognition of music emotion: A review.” *ACM Trans. Intell. Systems and Technology*, vol. 3, no. 3, 2012.

[14] R. E. Thayer, *The Biopsychology of Mood and Arousal*. Oxford University Press, Oxford, UK, 1989, ch. 2–5.

[15] T. Eerola and P. Toiviainen. *MIDI Toolbox: MATLAB Tools for Music Research.* University of Jyvaskyla, Finland, 2004. [Online]. Available: http://www.jyu.fi/musica/miditoolbox/

[16] C. Smit. *Midi Tools: Extending the midi toolbox in matlab*. Columbia University, New York, 2010. [Online]. Available: http://www.ee.columbia.edu/~csmit/matlab_midi.html

[17] M. Rohrmeier and I. Cross, “Statistical properties of tonal harmony in Bach’s chorales,” in *Proc. Int. Conf. Music Perception and Cognition*, Sapporo, Japan, pp. 619–627, 2008.

[18] *Songs with the same chord - Theorytab*. [Online]. Available: http://www.hooktheory.com/trends

[19] K. Lee, “Automatic chord recognition from audio using enhanced pitch class profile,” in *Proc. Int. Computer Music Conf.*, New Orleans, USA, pp. 306–313, 2006.

[20] T. Fujishima, “Realtime chord recognition of musical sound: A system using common lisp music,” in *Proc. Int. Computer Music Conf.*, Beijing, China, pp. 464–467, 1999.

[21] A. Metallinou and S. Narayanan, “Annotation and processing of continuous emotional attributes: Challenges and opportunities,” in *Int. Workshop on Emotion Representation, Analysis and Synthesis in Continuous Time and Space*, Shanghai, China, 2013.

[22] S. Mariooryad and C. Busso, “Analysis and compensation of the reaction lag of evaluators in continuous emotional annotations,” in *Proc. Humaine Assoc. Conf. Affective Comput. Intell. Interaction*, pp. 85–90, 2013.

[23] Y.-H. Yang and H. H. Chen, *Music Emotion Recognition*. Boca Raton, FL: CRC/Taylor & Francis, 2011, ch. 1–5.

[24] G. Kreutz, U. Ott, D. Teichmann, P. Osawa, and D. Vaitl, “Using music to induce emotions: Influences of musical preference and absorption.” *Psychology of Music*, vol. 36, no.1, pp. 101–126, 2007.