
Generation of Tempo Curve for Performance Rendering of Piano Music with Automatic Phrase Analysis

Naoya Ito
Satoru Fukayama
Daisuke Saito
Shigeki Sagayama

The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

NITOU@HIL.T.U-TOKYO.AC.JP
FUKAYAMA@HIL.T.U-TOKYO.AC.JP
DSAITO@HIL.T.U-TOKYO.AC.JP
SAGAYAMA@HIL.T.U-TOKYO.AC.JP

Abstract

This paper describes a method to generate tempo curves each of which is an envelope of tempo fluctuation, to realize more expressive performance. The purpose of performance rendering is to automatically generate expressive performance when a music score is given. In our approach to performance rendering, semi-global expressions which are realized by tempo control are focused on. As well as human performance, controlled tempo fluctuation reflects phrase structure of music, and it realizes a performance with high musicality. To achieve it, the proposed approach includes two method; automatic phrase analysis based on self similarity information of music and tempo curve generation by using the analyzed phrases. Evaluation results report that the method has a potential to generate expressive performances with high musicality.

1. Introduction

Performance rendering systems generate a human-like music performance from a music score. A large gap still exists between performance of human and that of computer. This is because human performance includes much expressions which are not directly written in score. Performance rendering systems should construct such expressions only from the scores.

Many machine learning methods for performance rendering systems have been proposed (Flossmann et al., 2009) (Tobudic & Widmer, 2003). Among them, (Kim et al., 2010) focused on *local expressions* which

In *5th International Workshop on Machine Learning and Music*, Edinburgh, Scotland, UK, 2012. Copyright 2012 by Naoya Ito/The University of Tokyo.

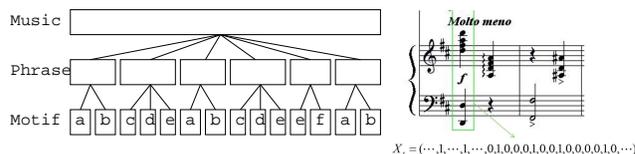


Figure 1. Hypothesis of music structure

Figure 2. Features

were realized by articulation, and articulation of each note was estimated by Conditional Random Fields (CRF) using attributes of the neighbor notes. Since articulation had an important role of human performance, the performance was generated with high human-likeness.

On the other hand, this paper focuses on *semi-global expressions* which are realized by tempo control. To realize it, interpretation of phrase structure of music is essential. Hence the generated performances are expected to be with high musicality.

By combining these rendering systems, more expressive performances can be generated with high human-likeness and musicality.

2. Automatic phrase analysis

2.1. Structural nature of music

To interpret music, a hypothesis about structural nature of music is built (See Fig. 1). It shows that: music consists of motifs and variants of them, and that several motifs construct a phrase. According to (Aiello & Sloboda, 1994), human express music performance with consciousness of some *units*, which may correspond to phrases in the hypothesis.

However, it is difficult to analyze a music structure because motifs cannot be defined directly unlike words in a sentence. Hence we focus on repeatedly used units and extract them as motifs.

2.2. Phrase extraction from a music score

In order to extract phrases, motifs should be extracted first. As mentioned above, motifs are defined as repeatedly used units. Hence, self-similarity matrix D_T of music is constructed as follows:

$$d(s, t) = D(X_s || X_t) = |X_s - X_t| \quad (1)$$

$$D_T = \begin{bmatrix} d(1, 1) & \cdots & d(1, T) \\ \vdots & \ddots & \vdots \\ d(T, 1) & \cdots & d(T, T) \end{bmatrix}. \quad (2)$$

In Equations (1) and (2), X_t is a vector denoting occurrences of MIDI note numbers, and $d(s, t)$ means hamming distance between X_s and X_t .

In matrix D_T , let us consider to find the optimal path from the first column to the T -th column with dynamic programming (DP). Finally motifs can be extracted as oblique lines in the optimal path. In this algorithm, there are three cases corresponding to local paths in DP; the case of progression in the same tempo, progression with a tempo fluctuation, and jump to another motif. The costs for the corresponding cases are set to α , β , γ respectively as $\alpha < \beta < \gamma$.

In addition, to deal with modulation correctly, additional similarity matrices composed by pitch-shifted features in column are prepared. Then DP is carried out using all the matrices. Finally, a phrase is constructed as an unit which is more than one measure following three measures by using some extracted motifs.

3. Generation of tempo curves

Several theories of music performance claim that ‘‘Pianists express a hierarchical structure in music with slowdown of tempo when a phrase finishes (Todd, 1985).’’, ‘‘Tempo accelerates when a phrase starts, tempo slows when a phrase finishes. A wide diversity of acceleration and slowdown exists.’’. According to them, tempo fluctuation has an important role to express phrase information. In our study, a tempo curve $f(t)$ is controlled by parameters (a_1, a_2, b_1, b_2) which represent depths and shapes of tempo acceleration and slowdown;

$$f(t|t_b \leq t \leq t_c) = -a_1 \left(\frac{2t - t_b - t_e}{t_b - t_e} \right)^{b_1} + 1 + a_1 \quad (3)$$

$$f(t|t_c \leq t \leq t_e) = -a_2 \left(\frac{2t - t_e - t_b}{t_e - t_b} \right)^{b_2} + 1 + a_2, \quad (4)$$

where t_b , t_e , and t_c are a start time of a phrase, an end time of a phrase, and an end time of acceleration, respectively. These parameters can be estimated using human performance data. In the proposed rendering system, these semi-global expressions and local expressions which are rendered by using conventional method are combined.

Table 1. Means and standard deviations of evaluation score

	Human-likeness	Musicality
w/o expressions	2.7±1.76	2.6±1.26
w/ local expressions	3.3±1.33	2.5±1.17
using proposed method	3.9±1.16	3.9±1.03
human performance	4.6±1.50	4.0±1.82

4. Experiments

4.1. Subjective evaluation

We used E.Grieg’s Arietta Lyric Piece Op.12-1 as a piece. We prepared 8 parameter sets for the proposed method which are manually selected. As reference, three additional sound samples are prepared; performance without expressions, performance with only local expressions by using ‘‘conventional’’ method (Polymymnia (Kim et al., 2010)) and human performance. Human-likeness and musicality of each sound sample were evaluated with 6 scaled scores by 10 participants, which consist of 5 non-pianists and 5 hobby-pianists who have experience of playing the piano over 5 years.

4.2. Results

The results of means and standard deviations of evaluation scores are shown in Table 1. It is observed that scores by using the proposed method were closer to those of human performance from both perspectives, especially in musicality. It indicates that more expressive performance is generated with high human-likeness and musicality, by combining local and semi-global expressions rendering systems. In addition, the proposed method has a potential to make performances more expressive with musicality.

5. Conclusion

In order to realize more expressive performance, we proposed a method to generate tempo curves with phrase analysis. The proposed approach includes two method; automatic phrase analysis based on self similarity information of music and tempo curve generation by using the analyzed phrases. Experimental evaluations show that the method has a potential to generate expressive performance with high musicality.

References

Aiello, R. and Sloboda, J. A. Music performance: Expression and the development of excellence. In *Musical Perceptions*, pp. 152–169. Oxford University Press, 1994.

Flossmann, S., Grachten, M., and Widmer, G. Expressive Performance Rendering: Introducing Performance Context. In *Proceedings of the 6th Sound and Music Computing Conference(SMC)*, pp. 155–160, July 2009.

Kim, T. H., Fukayama, S., Nishimoto, T., and Sagayama, S. Performance rendering for polyphonic piano music with a combination of probabilistic models for melody and harmony. In *Proceedings of the 7th Sound and Music Computing Conference(SMC)*, pp. 23–30, July 2010.

Tobudic, A. and Widmer, G. Relational IBL in music with a new structural similarity measure. In *Proceedings of the International Conference on Inductive Logic Programming*, pp. 365–382, July 2003.

Todd, N. P. A model of expressive timing in tonal music. *Music Perception*, 3(1):33–58, 1985.