9th International Conference on Music Perception and Cognition

Alma Mater Studiorum University of Bologna, August 22-26 2006

Pop-E: a performance rendering system for the ensemble music that considered group expression

Mitsuyo Hashida
Human Media Research Center,
Kwansei Gakuin University
Sanda, Hyogo, JAPAN
hashida@ksc.kwansei.ac.jp

Noriko Nagata
School of Science and Technology,
Kwansei Gakuin University
Sanda, Hyogo, JAPAN
nagata@ksc.kwansei.ac.jp

Haruhiro Katayose
School of Science and Technology,
Kwansei Gakuin University
Sanda, Hyogo, JAPAN
katayose@ksc.kwansei.ac.jp

ABSTRACT

This paper proposes a musical performance-rendering model, called Pop-E, for expressive ensembles: homophonic and polyphonic music. Polyphonic music has to be synchronized after the independent expression of each voice. Pop-E gives phrasing expressions to each part and has a means to align the timings of the parts according to the group structure and the "attentive" parts of the performance (attentive refers to parts that naturally draw the listeners attention). The paper describes how the middle part of a Pop-E performance of Chopin's "Fantasie-Impromptu" that received an award at NIME-Rencon was rendered. This paper also discusses the description ability of Pop-E, based on the reconstruction of three human performances of the piece. The experimental listening evaluation and reconstruction of the human performances suggested that Pop-E was able to render natural performances with a small set of performance rules.

Keywords

Musical group expression, performance rendering.

INTRODUCTION

Performance rendering is one of the main topics of musical information processing. Researches that ushered in performance rendering systems date back to the 1980's (Fryden et al. 1984, Clynes 1984). Since the 1990's, approaches involving music recognition theories such as Generative Theory of Tonal Music (Lerdahrl et al. 1983) or Implication-Realization Model (Narmour 1977), learning systems, and example base reasoning (Ishikawa 2002, Widmer 1996) have been proposed. In addition, a hearing competition for system-rendered performances has been held since 2002 (Hiraga 2002).

The quality of music performances generated by such systems has improved year by year since 2000; some rendering systems can now produce human-like performances of the main melody. Lately, the natural expression of multiple melodies has become an active topic (see Orchestra in a Box by Raphael (Raphael 2003) and iFP of Katayose (Katayose 2003)). However, these studies have been directed at interfaces to support musicians through the use of a piece as a template; they do not model methods of expression for performing music with multiple parts.

We aim to design a basic rendering model and its interface to facilitate natural expression of music comprised of multiple parts. This paper describes present state of performance rendering in section 2. Sections 3 and 4 describe the design of the model and the prototype system. Section 5 describes evaluations of our model and performances rendered with it.

PRESENT STATE OF PERFORMANCE RENDERING

Computer performance rendering to realize expressive performances from an input score is an active topic in the domain of information processing. In the 1980's, Sundberg et al. devised performance expression rules for chords, phrase changes, etc. by conducting analysis-by-synthesis studies of musicians and developed a performance system using those rules (Friberg 1984). Clynes developed an expressive system that used combination of tempo and velocity ratios in two or three events of a hierarchical metrical structure (Clynes 1984).

Since the 1990s, several studies have appeared on autonomic systems that render expressions by not only applying rules. Widmer extracted performance rules through a combination of inductive learning (IBL-SMART) and interpolation (Widmer 1996). Then he focused on constitution of a performance model that generates phrasing expression by searching for similar phrases with the nearest neighbor method (Widmer 2002). Bresin et al. improved their system “Director Musices”, and discussed the propriety of performance rules on emotional expression of articulation (e.g., *legato, staccato*) on each note by comparing a rule-based system with a neural network system (Bresin 1996). Katayose et al. worked on system learning and extraction of performance rules. They extracted common features of expression in a melody pattern (Katayose, 1990) and extended their method to iteration of multiple regression analysis (Ishikawa 2002). They also worked on phrase analysis based on a corpus (Katayose 2002).

Since 2000, performance systems have grown in their capability of rendering, and their expressiveness for certain performances has earned good opinions from audiences; they have been rated as equal to performances by human players. Lately, interest has risen in the expression of performances comprising multiple parts. Systems such as “Orchestra in a Box” (Raphael 2003) and iFP (Okudaira 2003) aim to be interfaces for human players—they provide means to control dynamics and tempo— and use human performances as templates for the expression of multiple parts.

There are three problems with the present systems as follows.

Automation of Group Structure Analysis

It is difficult to identify a certain group structure of a piece that has a multiplicity of interpretations. Although it is clear that recognition music theories such as GTTM have contributed to the solution of this problem, there remains the problem of changing priority among preference rules.

Formalization of Performance Expression

Many performance rules based on a musician’s rule of thumb have been suggested. However, performance expression techniques have yet to be formalized. In particular, there is still no performance model applicable to an affective piece such as Chopin's “Chanson de l’adieu” of which almost all current systems express poorly.

Development of Performance Rendering Model for Ensembles

A recent rendering study suggested a system that assumes natural expression as the principal objective of a number of melodies. Some performances of these systems received good evaluations from audiences at the *Rencon* contest (Hiraga 2002); however, these systems don’t use a common model but instead use a human performance as a template. Relative to the above subject, a new rendering system is required to include an interpretation model for multi-part music.

We developed a performance rendering model for multi-part music. The system embodying the model can be used to improve the quality of the rendered performance and reduce the user's burden of analysis.

Pop-E: PERFORMANCE EXPRESSION MODEL FOR *POLYPHRASE* ENSEMBLE

We propose a performance expression model for multipart music, Pop-E (Polyphrase Ensemble), to (1) render a natural expression of multipart music and (2) support a user so that he or she can efficiently design a performance. In Pop-E, an independent group structure is initially given to each part, after which two kinds of performance rule (group expression, expression of temporal note stretching) are applied to the parts. Through this processing, a temporal gap occurs in the occupancy time, which is necessary for a performance with numerous parts. This problem is resolved by the introduction of an *attentive part* (defined below) and a new function of synchronization between parts. For the second aim, we plan a formulation of performance rendering processing and promotion of the design of that processing assuming human manipulation.

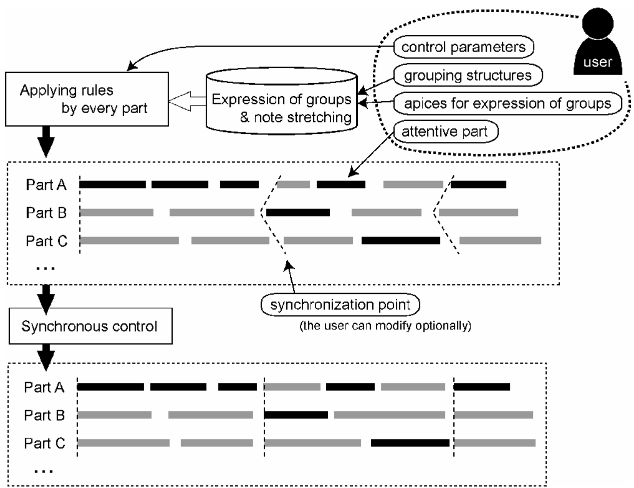

Figure 1. Concept of Pop-E.

ISBN 88-7395-155-4 © 2006 ICMPC

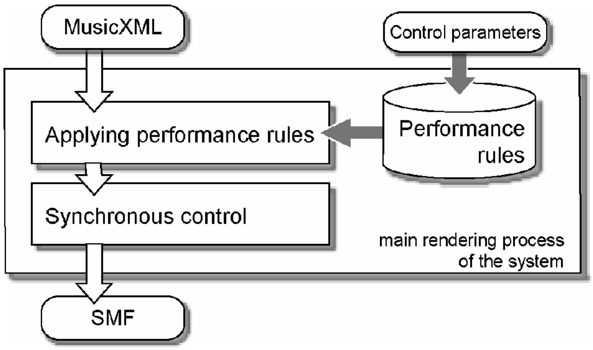

Figure 2. Outline of rendering processing**.**

Figure 1 illustrates the concept of our performance model, and Figure 2 is an outline of the performance rendering process. The user inputs the group structure (a pair comprising the beginning and the last notes of a group) of each part, the apex note of the rainbow-law expression of each group, the *attentive part*, and performance control parameters. It is assumed that each part consists of monody. The group structure of each voice is utilized when the group expression rules are applied and candidate positions are estimated in synchronous processing between parts. The *attentive part* is utilized as the criteria of the synchronous control and for note stretching. Next, we describe the music structures, performance rules, and the concept behind the rendering process.

Group Structure

A group is a sequence of adjacent notes that constitutes a musical unit such as a phrase. A group is placed hierarchically by voice. Voice parts are prohibited from having a nested structure.

Pop-E uses a pair of notes (beginning note, last note) to indicate each group. All the notes constitute a group on the same part, and all notes belong to an either group by all means. In adjacent groups, the beginning note of the preceding group and the last note of the following group can overlap considering the intent of the composition. For the same reason, we assume a group structure consists of independent parts, and the structures of two parts may not always be the same.

It is difficult to estimate a certain group from a score exactly. As such, we make the user input a group structure based on three criterions - beams, slurs and the grouping preferences for local levels on GTTM (Lerdahl et al. 1983). These inputs enable the interface to deal with variable interpretation.

Attentive Part

The attentive part is a sequence of intermittently remarkable sounds in a piece that occur regardless of the role of the voice part. An example of an attentive part is shown in Figure 3. Basically, it is a sequence of notes that are easily remembered and sung to oneself. Although, there is as yet no formal procedure to identify such parts, we note that such identification is natural and that identification of these parts would require little effort from users.

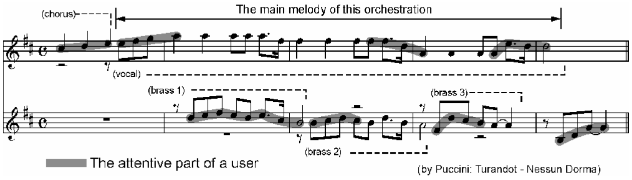

Figure 3. An example of the attentive part of a user.

Performance Rules

In Pop-E, a performance is rendered by changing the dynamics and tempo, and this is accomplished by sequentially applying rules to a note to which the input conditions match. The control objects are the velocity (dynamics) and inter-onset interval (IOI) in MIDI format (tempo) [1]. Our policy is to guarantee the ease of use of the performance rendering system interface and to construct the interface so that it works well with fewer rules. Here, we use just two methods, the expression of groups and temporal note stretching.

Expression of groups

The expression for group structure is a fundamental issue in music performance. Similar to other performance systems (Ishikawa 2002), in Pop-E, two kinds of group expression can be managed as representative expressions.

(a) Giving an accent of velocity to the beginning note of a group. Controlling the dynamics of sound production is one of the basic techniques for handling an instrument. In Pop-E, the beginning note is weighted with the initial velocity.

(b) *Rainbow-law* expression. This expression increases the dynamics and tempo from the beginning note to the apex note of a group, after which it decreases them to the last note of that group to become the climax of that expression at the apex note (see Figure 4). This expression, which is given to each group, is one of the most important for the musical representation of a phrase (Hoshina 1998). The Pop-E procedure is to indicate the apex note of a group and then starting with the initial velocity and length, to increase the values of the other notes until they reach the apex values and after that to decrease them linearly to the initial values. The rule to increase and decrease the velocity and length has a past successful record in a rendering system (Ishikawa 2002). The rainbow-law expression can be applied to any level of a group structure.

[1] The offset-time of a note is another important parameter for performance handling. It has the strong relationship with articulation. However, we think it that it would make sense to include it in a model for pedaling, which is an important topic in performance rendering.

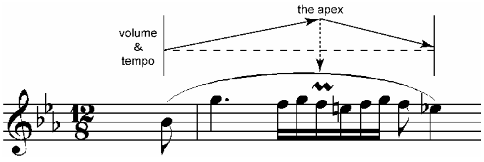

Figure 4. Example of *rainbow*-law expression.

Expression of temporal note stretching (agogik)
Another important factor for expressive performances is *agogik*: a small change of tempo or a mute or expansion of a sounding note. We are concerned with stretching the length of a note. Our model deals with the following notes.

(a) Grace notes and k-level tuplets. The ratios of the initial values of velocity and length to all notes to accord are given. K-level tuplets are used when notes appear irregularly rather than periodically as a rhythm pattern.

(b) Leaping notes. When the interval of pitch of two adjacent notes in the same part is larger, the ratios of initial values of velocity and length to which other values are given to the precedent note.

(c) Attention transition notes. When the attentive part steps over other parts, the ratios of initial values of velocity and length to which other values are given to notes on the boundary. Each parameter of these ratios can be given by the hierarchical level of the grouping structure or combination of parts. An example of attention transition notes is shown in Figure 5.

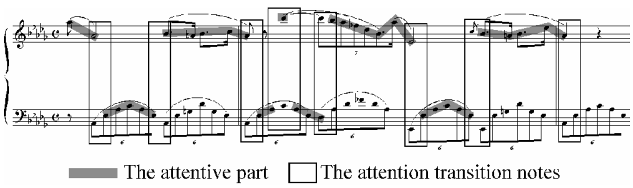

Figure 5. Attention transition notes.

(a) and (b) are gotten from scores, and (c) is gotten from the part information of the score and the attentive part.

Synchronous Control

In Pop-E, the performance rules are applied to each component without specifying the musical role (melody, base, accompaniment, etc.). This procedure causes an unexpected timing lag between some parts in the same area of the score; the performance would not make any sense if that lag is neglected. To solve the problem and keep the individual expression of the parts, we need to estimate synchronization points to align parts. We then have to schedule the timing of all parts according to the timing of the attentive part.

Estimation of synchronization point
The synchronization points are demanded by comparing group structure given to each part each other. These points indicate the onset times of notes of the beginning or the last note in any group. The notes at the synchronization point sound at the same time.

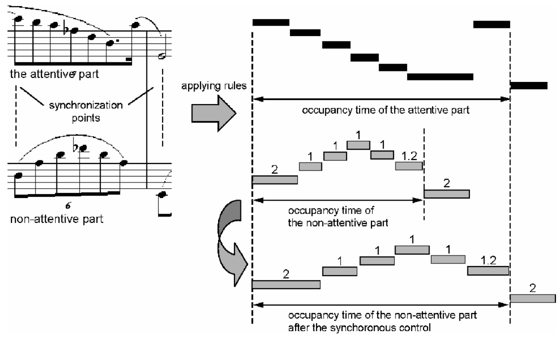

Figure 6. Scaling occupancy time of non-attentive parts.

Scaling occupancy time of non-attentive parts
In the area between two adjacent synchronization points, the relevant non-attentive parts are scaled linearly while maintaining the note length ratio. For example, if the upper part of Figure 6 is the attentive part, the lower part becomes the objective to be scaled. The right side of Figure 6 shows the length of each note, and their numbers are the length ratios of the performance rule.

IMPLEMENTATION OF RENDERING PROCESS

This section describes the implementation of the prototype system based on Pop-E. The system basically inputs score information (pitch, duration and onset time) and the music structures described in section 3, puts performance information (velocity and length) on the notes that are consistent with the performance rule, and outputs performance data in SMF format through the synchronous control.

The main unit of performance rendering processing of Pop-E is implemented on OPS/R2 for the forward reasoning engine, and consists of performance rules and a control module. The system reads a score in MusicXML format and control parameters as external information. It allocates this information to working memory, applies performance rules, executes synchronous control, and outputs expressive performance data in SMF format. This output data is converted into an audio signal by a commercial synthesizer.

Input Format

The inputs into the main processing for performance rendering are basic score information (pitch, duration and onset time), group structure (beginning and end notes of a group), the apex notes for the rainbow-law expression of every group, and the attentive part. These inputs are written in MusicXML format in consideration of software compatibility.

The initial value of onset timing of each note t_n^0 is given by adding the onset timing and duration of the preceding note as follows:

ISBN 88-7395-155-4

$$t_n^0 = t_{n-1}^0 + d_{n-1} \text{ (if } n\text{=0 then } t_n^0 = 0)$$

Applying Performance Rules

The control parameters for rendering performances are velocity (v) and length (l) of each note (n). Notes include rests. Our system deals with the time length (IOI) possessing the onset time at both ends of two adjacent notes in the same part as the basic control object of tempo expression. The onset time t of each note is expressed as follows:

$$t_n = t_{n-1} + l_{n-1} \text{ (if } n = 1 \text{ then } t = 0)$$

In the rendering processing, knowledge of performance expression, namely a performance rule $R_i(v_n, l_n)$, is applied to v and l of each note. In each performance rule, control parameters p_v and p_l (ratio for initial value v_n^0 and l_n^0) are multiplied with the velocity and length. p_v and p_l are described as external data in order to correspond to a rule.

$$R_i(v_n, l_n) = \{p_v \times v_n^0, p_l \times l_n^0\}$$

A performance expression $E(v_n, l_n)$ for note n is expressed as follows.

$$E(v_n, l_n) = \prod_{i=1}^{r} R_i(v_n, l_n)$$

In the rules of rainbow-law expression, v_n and l_n are modified into v_k and l_k, which are calculated as follows.

$$v_k = v_n \times \frac{Time_{G_k}}{Time_G}, \quad l_k = l_n \times \frac{Time_{G_k}}{Time_G}$$

$Time_G$ expresses the occupancy time of the present group, and $Time_{Gk}$ expresses that of the k-th note among that group. These parameters are expressed as follows:

$$Time_G = \sum_{i=0}^{m} t_i \text{ (} m\text{: the last note of a group)}$$

$$Time_{G_k} = \sum_{i=0}^{k} t_i \ (k < m)$$

Synchronous Control

The synchronous control aligns parts with matching onset timings on synchronization points. The basic strategy is to expand or contract occupancy time of non-attentive parts linearly along the attentive part between synchronization points. An example is shown in Figure 7. The synchronization points are marked with *t and the scopes of scaling are enclosed with rectangles. The synchronous control deals with two parts, the attentive part and a non-attentive part. First, the scopes of both parts (T_{att}, T_{non}) are calculated by adding the lengths of the notes of each part. Then the scope of the non-attentive part is multiplied by T_{non}/T_{att}. This procedure is recursively executed until all synchronization points of every part match.

Figure 7. Scopes of synchronous control.

EVALUATION

We created Pop-E in consideration of the quality of the rendered performance and reduction of the user's burden. The model's design depended on our naive opinions about music. We thus had to evaluate our model as objectively as possible. This section describes two evaluations, a hearing examination conducted at NIME-Rencon and a rendering examination of reconstructed human performances.

Hearing Evaluation

We contributed a performance rendered by Pop-E to NIME-Rencon [2] (Noike 2005). There, the performance scored the highest of all the systems and won the Rencon Award. Below, we give an overview of NIME-Rencon and Pop-E's performance of the middle part of Chopin's "Fantasie-Impromptu."

Content of NIME-Rencon and evaluation method

The hearing contest of NIME-Rencon featured performances of five systems and two human players. The audience wasn't given any information on who or what was performing (e.g., what system rendered it). The method of evaluation was five-step voting regarding (1) human likeness and (2) musical preference. The scores of each system were the averages of all votes or the averages of (2) if two or more systems got the same value. The performances were ranked by their score, and the system given the highest score won the Rencon Award. The valid votes totaled fifty-one.

Performance rendered by Pop-E

The performance data of the middle part of Chopin's "Fantasie-Impromptu" (see Figure 8) consisted of two parts. There are 37 pairs of all 204 notes of both parts are accorded to the onset time. The note sequence of the lower part consisted of sextuplets, and that of the higher part included some triplets and complicated septuplets combined

[2] Rencon is a performance rendering contest, which aims to build a basis for objective evaluation of music systems. It was started in 2002. (http://shouchan.ei.tuat.ac.jp/~rencon/)

ISBN 88-7395-155-4

with a tie from the preceding note. Both parts must be expressed delicately, especially the tempo in the second measure. Thus, we thought this piece would be applicable to investigate the abilities of Pop-E.

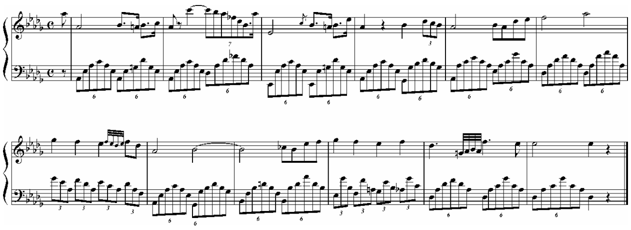

Figure 8. The middle of Chopin's "Fantasie-Impromptu"

The performance rules were restricted to the following nine rules for simplicity: giving an accent of velocity to the beginning note of a group (two parts), the rainbow-law expression (two levels) grace notes (either part), sextuplets (the other part), leaping notes (upper and lower directions) and the attention transition notes (two levels)[3]. A performance based on a performance control parameter determined from experience was generated.

Table1: Control parameters for "Fantasie-Impromptu" (upper part / lower part)

Rule	Velocity	Tempo
Initial value	64/48	-
Accent to the beginning of a group	1.3/1.5	1.0
Rainbow-law (Lv.1)	1.25	1.0
Rainbow-law (Lv.2)	1.5	1.0
Grace notes & k-tuplets	1.0	2.0
Attentive parts transition	1.0	2.0/1.25
Leaping notes (upper)	1.0	1.5
Leaping notes (lower)	1.0	2.5

Table 1 lists the performance control parameters. The fractional values show magnification of the initial value; 1.0 means not to vary velocity (or length). The initial tempo in BPMs (beats per minute) was 112 quarter notes. The duration of the grace notes on the upper part, which did not describe information in MusicXML[4], was equal to a thirty-second note (about 67 milliseconds) for descriptive purposes. The performance data can be listened to at http://www.m-use.net/research/PopE/.

Results

Figure 9 shows the voting results of the human-likeness test. The two human performances and the Pop-E performance were given high scores. There was a significant difference with a 1% standard between the Pop-E performance and the other system performances. The results of performance preference test were similar to those of human-likeness. Consequently, Pop-E's performance won the Rencon Award.

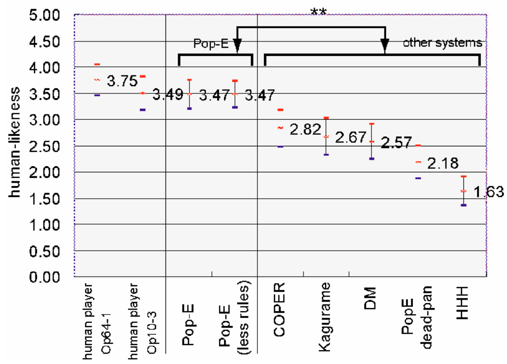

Figure 9. Results of human-likeness test at NIME-Rencon.

Discussion of Reconstruction Capability

There is another method of evaluating performance expression models, which is to value a performance reconstructed by a set of control parameters tuned for a certain performance in the system (Clynes 1984, Ishikawa 2002). Here, we reconstructed three performances by three pianists (Heming, F., Nakamura, H. and Ashkenazy, V.) and evaluated them with a human hearing test and a quantitative comparison of correlation factors between the reconstructed and the objective performances. The control parameters that were set on the reconstruction are shown in Table 2 (see the last page of this paper). The procedure to generate the reconstruction performances is: (1) make a target performance by manually translating the sound signal of a player's actual performance into MIDI data, (2) set control parameters of which correlation factors of velocity and length between the target performance and the reconstructed one become greatest, and (3) generate the reconstructed performance with the parameters.

Figure 10 shows the results of the reconstruction by Pop-E, and Tables 3 to 5 lists the correlation factors between the reconstructed performance and the target performance and the other target performances.

[3] In addition, a simple rule for pedaling was devised; the pedal information (the control change 64 (sustain) in MIDI message) was given linearly from 0 to 127 within the length of each note at the base line of the lower part.

[4] MusicXML is a universal translator for common Western musical notation from the 17th century onwards. It is developed by Recordale LLC (http://www.recordare.com/xml.html)

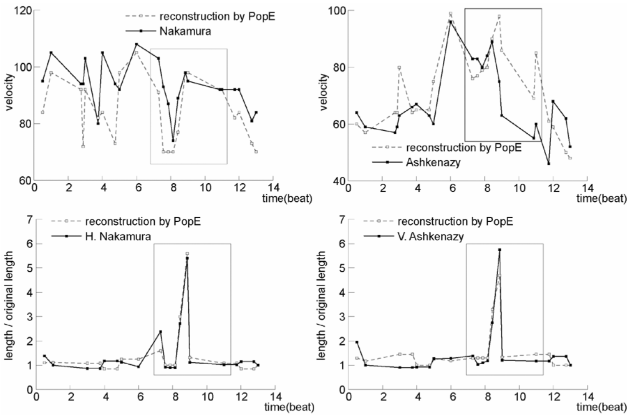

Figure 10. Reconstructions of three players

Table 2: Reconstructions of three players

	Heming		Nakamura		Ashkenazy	
	velocity	tempo	velocity	tempo	velocity	tempo
initial value	40/38	-	70/45	-	48/35	-
accent to the beginning of a group	1.2/1.1	0.8/1.2	1.2/1.2	1.1/1.2	1.25/1.1	1.3/1.0
rainbow-law (Lv.1)	1.5/1.0	1.0/0.8	1.2/1.0	0.8/1.0	1.1/1.2	1.0/1.3
rainbow-law (Lv.2)	1.5/1.3	1.0/1.0	1.0/2.0	1.3/1.4	1.7/1.8	1.0/1.7
grace notes	1.0/--	1.0/--	1.0/--	1.0/--	1.0/--	1.3/--
k-level tuplets	1.1/--	0.7/--	0.9/--	1.3/--	1.1/--	0.7/--
attention transition (type 1)	1.8/1.2	1.1/1.2	1.4/1.1	1.0/1.0	1.2/1.3	1.1/1.0
attention transition (type 2)	1.3/1.1	1.5/1.1	1.3/1.1	1.6/1.5	1.2/1.1	1.3/2.0
leaping notes (upper)	1.0/1.1	2.7/1.3	1.1/1.0	2.7/1.0	1.2/1.0	3.0/1.0
leaping notes (lower)	1.0/0.9	3.4/2.0	1.2/1.0	5.5/2.2	1.2/1.0	4.0/2.0

Table 3: Correlation factors between target (all players) and reconstruction (velocity/tempo)

Part	Same player	Other players
Upper	0.70/0.91	0.37/0.86
Lower	0.53/0.68	0.41/0.38
Average	0.59/0.76	0.40/0.55

Table 4: Correlation factors between target (Nakamura) and reconstruction (velocity/tempo)

Part	Nakamura	Heming	Ashkenazy
Upper	0.65/0.97	0.6./0.79	0.00/0.97
Lower	0.36/0.83	0.44/0.17	0.68/0.49
Average	0.48/0.89	0.50/0.39	0.44/0.66

Table 5: Correlation factors between target (Ashkenazy) and reconstruction (velocity/tempo)

Part	Ashkenazy	Heming	Nakamura
Upper	0.66/0.94	0.41/0.78	0.28/0.91
Lower	0.70/0.69	0.42/0.30	0.20/0.43
Average	0.69/0.77	0.23/0.60	0.48/0.72

In the comparative hearing evaluation, the listener needed to have the ability to understand the expressive details in a performance. Therefore, this examination was done with ten listeners who have experience playing musical instruments or who listen regularly to classical Western music. They were given the information about the three players of the target performances beforehand. We told them to listen to the targets and the reconstructions repeatedly, and asked them (1) if they found any difference between the performances, and (2) which performances corresponded to each other.

As a result, all subjects heard differences between the target performances and could specify the correspondence with the target and the reconstruction. Some subjects pointed out that the tempo features of each player appeared at the sep-tuplets and lasted until the beginning of the third measure (the rectangles in Figure 10) in the reconstructions. As shown in Tables 3 to 5, the average correlation factor between the target and the reconstruction of the same target player was 0.59 for velocity and 0.76 for tempo. On the other hand, the correlation to a different player was 0.40 for velocity and 0.55 for tempo. In addition, the quality of reconstruction was (a) better in regard to the velocity than to tempo and (b) better in the upper part than in the lower part.

Regarding the correlation factors by player, the correlation factor of tempo was 0.89 for Nakamura, and the correlations were 0.69 for velocity and 0.77 for the tempo for Ashkenazy. Ishikawa et al. reported that their experimental results (0.655 for velocity and 0.273 for tempo) corresponded with those of our test, by fitting of the least squares approximation using 21 rules (Ishikawa 2002). The aim of their system is precision enhancement of the reconstruction through the addition of performance rules, using AND processing by iteration of the multiple regression analysis. They didn't focus on the design of a fundamental performance rule. Their results cannot easily be compared with ours because the object music was different. Despite this, we can say that the precision of Pop-E, which used nine rules, exceeded the precision of Ishikawa's system. We think that this result shows that Pop-E could capture the essence of performance interpretation.

In Figure 10, the correlation factor for velocity is not higher than the one for length. This result tells that there is a limit to reproducing accurate expressions with only nine control parameters. The hierarchical level to give rainbow-law expression was limited to two phases; the correlation factors could be increased if these levels were enhanced. The listeners were able to distinguish performances with limited parameters, and they seemed to do so by using length, i.e., tempo expression, as a clue. To confirm this, we rendered one-hundred performances by interpolating parameters between those of Ashkenazy and Heming and did a discrimination test with four listeners (three experienced listeners and a non-experienced listener). As a result, decision boundaries of all listeners consisting of tempo predominance were confirmed.

CONCLUSION

This paper presented a performance-rendering model (Pop-E) for music consisting of multiple parts. The aim of this research is to develop a general means of expression and to provide efficient support for performance design.

A user of this system has to input structural information such as the group structure, apex note, and attentive part. The system enables the user to prepare performance data as a whole, because the performance rules were carefully selected and many other procedures for performance rendering were implemented in the model.

The performances generated by Pop-E using five rules (nine parameters), which are fewer for rule-based rendering systems, won the Rencon Award. In addition, Pop-E could have listeners to hear a difference of three performance expression. We think that Pop-E could consider one end of essence of performance expression.

As future topics of study, we will use pedaling to handle the offset timing of notes and design an interface that gives a music structure as a support for the user.

ACKNOWLEDGMENTS

This research was partially supported by PRESTO “Intelligent Cooperation and Control” JST and CREST “Foundation of technology supporting the creation of digital media contents”, JST, JAPAN.

REFERENCES

Fryden, L. and Sundberg, J. (1984). Performance Rules for Melodies. Origin, Functions, Purposes. *Proc. Of International Computer Music Conference (ICMC),* pp. 221-225.

Clynes, M.(1984). A Composing program incorporating Microstructure. *Proc. of International Computer Music Conference (ICMC).* pp. 225-232.

Lerdahl, F. & Jackendoff, R. (1983). *A Generative Theory of Tonal Music.* MIT Press.

Narmour, E (1977). *The Analysis And CognitionOf Basic Melodic Structures*, the University of Chicago Press

Ishikawa, O., Katayose, H., & Inokuchi, S. (2002). Identification of Music Performance Rules Based on Iterated Multiple Regression Analysis. *Journal of IPSJ, Vol. 43, No. 2,* pp. 268-276. (written in Japanese)

Widmer, G. (1996). Learning Expressive Performance: The Structure-Level Approach. *Journalof New Music Research, Vol. 25, No. 2,* pp. 179-205.

Hiraga, R., Hashida, M., Hirata, K., Katayose, K. & Noike, K. (2002). RENCON: Toward a New Evaluation Method for Performance Rendering System. *Proc. of International Computer Music Conference (ICMC).* pp. 357-360.

Arcos, J., de Mantaras, R. & Serra, X. (1998). SaxEx: A Case-Based Reasoning System for Generating Expressive Musical Performances. *Journal of New Music Research, Vol. 27, No. 3,* pp. 194-210.

Hiraga, R., Hashida, M., Hirata, K., Katayose, H. & Noike, K. (2002). Rencon: toward a common basis for performance rendering concours, *Proceedings of ICAD 2002 Rencon Workshop*, pp. 51-53.

Raphael, C. (2003). Orchestra in a Box: A Systemfor Real-Time Musical Accompaniment. *International Joint Conference of Artificial Intelligent (IJCAI): workshop program APR-5,* pp. 5-10.

Katayose, H. & Okudaira, K. (2004). iFP A Music Interface Using an Expressive Performance Template, *Entertainment Computing 2004, Lecture Notes in Computer Science, Vol. 3166,* pp. 529-540.

Widmer, G. & Tobudic, A. (2002). Playing Mozart by Analogy: Learning Phrase-level Timing and Dynamics Strategies, *Proc. of International Conferenceon Auditory Display (ICAD),* pp. 28-35.

Bresin, R. & Battel, G. (2000). Articulation strategies in expressive piano performance: Analysis of legato, staccato, and repeated notes in performances of the andante movement of Mozart’s Sonata in G major (k 545), *Journal of New Music Research, Vol. 29, No. 3,* pp. 211-224.

Katayose, H. & Inokuchi, S. (1990). Kansei Music System, *Computer Music Journal,Vol. 13, No. 4,* pp. 72-77.

Katayose, H., Uwabu, Y. & Ishikawa, O. (2002). A Music Interpretation System: Schema aquisitionand Peroformance Rule Extraction, *Proc. of ICAD2002 Rencon Workshop,* pp. 7-12.

Hoshina, H. (1998). *The Approach toward a Live Musical Expression: A Method of Performance Interpretation considered with energy,* Ongaku-no-tomo-sha. (written in Japanese)

Noike, K., Hashida, M., Hirata, K., Katayose, H. & Hiraga, R. (2005). A Report of NIME04 Rencon and Next Plan, *IPSJ-SIG, 2005-MUS-59,* pp. 71–76. (written in Japanese)