Laminae: A stochastic modeling-based autonomous performance rendering system that elucidates performer characteristics

Kenta Okumura
Nagoya Institute of Technology
k09@mmsp.nitech.ac.jp

Shinji Sako
Nagoya Institute of Technology
sako@mmsp.nitech.ac.jp

Tadashi Kitamura
Nagoya Institute of Technology
kitamura@nitech.ac.jp

ABSTRACT

This paper proposes a system for performance rendering of keyboard instruments. The goal is fully autonomous rendition of a performance with musical smoothness without losing any of the characteristics of the actual performer. The system is based on a method that systematizes combinations of constraints and thereby elucidates the rendering process of the performer's performance by defining stochastic models that associate artistic deviations observed in a performance with the contextual information notated in its musical score. The proposed system can be used to search for a sequence of optimum cases from the combination of all existing cases of the existing performance observed to render an unseen performance efficiently. Evaluations conducted indicate that musical features expected in existing performances are transcribed appropriately in the performances rendered by the system. The evaluations also demonstrate that the system is able to render performances with natural expressions stably, even for compositions with unconventional styles. Consequently, performances rendered via the proposed system have won first prize in the autonomous section of a performance rendering contest for computer systems.

1. INTRODUCTION

In recent times, several autonomous systems for automatic performance rendering have been proposed [1, 2]. Their main motivation is elucidation of the existing performance and the realization of a virtual performer [3, 4]. Such systems generally control the rules that determine performance expression without asking for interaction with the user in the rendering process of the performance. Our focus is on the ability to render performances without losing any of the characteristics of the human performer, and to replicate such characteristics. One of the most rational ideas for achieving this is to relate the expression included in segmented cases of the performance of human virtuosi and the information that describes the conditions in which they were performed.

The typical method used to handle expressions included in each case is to transcribe the statistical trend in sections of accumulated cases [5–7]. The advantage of that method is that unnatural expressions are less likely to occur in the rendered performance. However, that method is not necessarily advantageous as it may not faithfully reproduce the performer's characteristics, since the features of the performer that were originally provided in the cases are smoothed by the statistics. Conversely, there is a method that directly transcribes the expression of the particular case among the cases that have been accumulated [8–10]. This is a more suitable method for faithful reproduction of the performer's characteristics because of its certain retention of the feature of the cases. However, the problem with this method is that the performances may lose naturalness since they are rendered by connecting cases that are not continuous in the original performance. In the existing methods, the rules used to select the case are not optimized for the composition to render a performance by the system because they are constructed based on the compositions originally performed by the performer. To solve this problem, we propose a method that searches for the optimum case to transcribe the expression from the alternatives, augmented by the moderation of a strict rule. This is done with the assumption that the possibility exists a case with an expression that can render a more natural performance exists in those cases that are never selected because they are not strictly in accordance with the selection rule.

The information that describes the conditions of the case that was originally performed must be elucidated with generality to select the optimum case for every direction upon rendering of the performance. Most existing autonomous systems require the information related to the interpretation of the composition by the performer as input. However, it is difficult to acquire rules that can accurately describe the relationship, even when it is analyzed by experts. In addition, to explain the relationship with generality is also difficult for the performer because of fluctuations in the interpretation itself [11]. We consider an approximate description of the relationship using the combination of simple information obtained uniquely from the score rather than a higher-order interpretation of the performer. We previously proposed a method that enables systematic association of the relationships without using such unstable information, under the assumption that there is a tendency in the behavior of the performer that depends on the context of the performance directions locally derivable from the score [12]. That method is able to eliminate the dependency on any information other than the performance itself, because it uses no such information containing the fluctuations mentioned above. The essence of the problem that the method resolves, in terms of classifying the cases of existing performances based on the information from the score, is congruent with our proposal.

2. METHODS CONSTITUTING THE SYSTEM

Performers interpret the directions $S = (s_1, \ldots, s_M)$ that are notated in the given score, and renders the performance sequence $\hat{R} = (\hat{r}_1, \ldots, \hat{r}_N)$ by applying their intended expression. On the assumption of the sequence of strict direction $\hat{S} = (\hat{s}_1, \ldots, \hat{s}_N)$ that represents the contents of the performance, the applied expressions are observed as sequences $D = (d_1, \ldots, d_N)$ for factors $F = (\mathrm{AT}, \mathrm{GR}, \mathrm{DR}, \mathrm{BR})$ between $\hat{S}$ and $\hat{R}$ as follows:

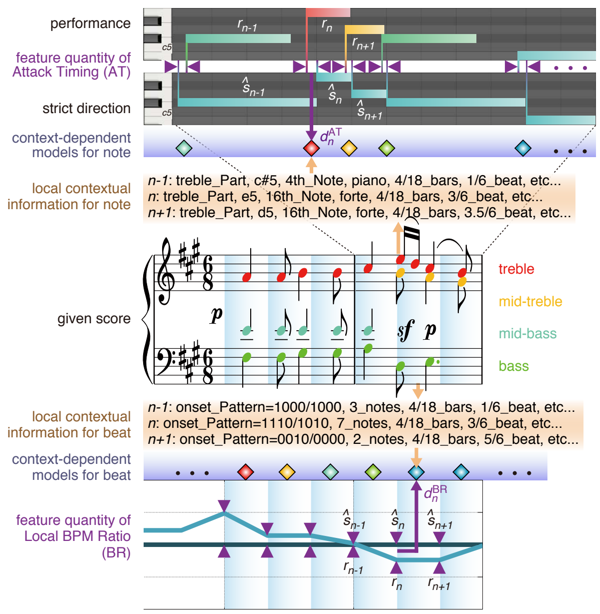


Figure 1: Formation of context-dependent models.

D^{AT} (Attack Timing): Timing of striking of the key in beats per quarter note.

D^{GR} (Gatetime Ratio): The ratio of the time taken to depress a key in the performance to that note's length on the score. If the length of the performance is shorter than the score's instruction, the value is less than one.

D^{DR} (Dynamics Ratio): Dynamics of keying in the ratio of the notated dynamics. The value is acquired in the same manner as D^{GR}.

D^{BR} (Local BPM Ratio): Ratio of the beat's BPM to the average BPM of the performance.

These are the main ingredients of the performance expression that are utilized in the operation of the instrument under the artistic intention and physical constraints of the performer [13]. We also observe the difference in their quantities between the preceding feature quantities, since it is believed that the rendering of various quantities depend on the tendency of their preceding direction. In the case of performance $\hat{r}_n$ and its direction $\hat{s}_n$, the feature quantities and such differences for the factors F are extracted by the following equations:

$$d_n^F = \begin{cases} \hat{r}_n^F - \hat{s}_n^F, & F = \mathrm{AT} \\ \hat{r}_n^F / \hat{s}_n^F, & F = (\mathrm{GR}, \mathrm{DR}, \mathrm{BR}) \end{cases}, \quad (1)$$

$$d_n^{\Delta F} = d_n^F - d_{n-1}^F, \quad F = (\mathrm{AT}, \mathrm{GR}, \mathrm{DR}, \mathrm{BR}) \ . \quad (2)$$

Even in the performance based on the score, another series of cases is excited if a trigger note that has the direction of insertion of notes, such as *trill*, for example, exists in the vicinity. The following sequences of information $X = (x_1, \ldots, x_N)$ are described to consider the general possibility that the number of cases for the note is $M \simeq N$:

X^{PS} (Pitch Shift): Integer value of the distance from the pitch directed by the score. The value is usually zero.

X^{KS} (Key Strokes): Number of notes performed for the corresponding note in the score. The value is usually one.

This information makes possible to associate plural cases for performance direction s_m. The system can render the

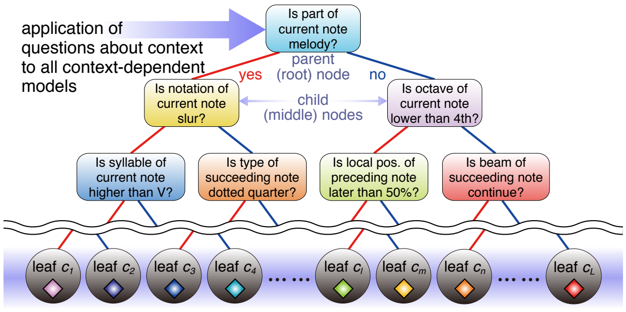


Figure 2: Systematization of context-dependent models.

performance sequence $\hat{V} = (\hat{v}_1, \ldots, \hat{v}_N)$ that accommodates the possibility of such a mismatch by referring to information in x_m corresponding to v_m, if the optimum series of cases $V = (v_1, \ldots, v_M)$ to perform the sequence of score S is determined by searching for cases that qualify as candidates using the method discussed later.

2.1 Modeling and systematization of the cases

In this proposal, cases from existing performances are made selectable by using only the performance direction information available from the score. Context-dependent models for each case are defined to describe the relationship of feature quantities and strict direction (Figure 1). The tendency of G factors of feature quantities and difference in the case of $\hat{r}_n$ based on $\hat{s}_n$ are regarded in this model as the multivariate normal distribution with the probability density function shown in the following equation:

$$P\left(\boldsymbol{d}_n | \boldsymbol{\mu}_n, \boldsymbol{\sigma}_n\right) = \prod_{f \in F} P\left(d_n^f | \mu_n^f, \sigma_n^f\right) = \frac{\exp\left\{-\sum_{f \in F} \frac{\left(d_n^f - \mu_n^f\right)^2}{2\sigma_n^f}\right\}}{\sqrt{(2\pi)^G \prod_{f \in F} \sigma_n^f}} \quad (3)$$

$$\begin{cases} F = (\mathrm{AT}, \mathrm{GR}, \mathrm{DR}, \Delta\mathrm{AT}, \Delta\mathrm{GR}, \Delta\mathrm{DR}), & G = 6, \quad \text{for note} \\ F = (\mathrm{BR}, \Delta\mathrm{BR}), & G = 2, \quad \text{for beat} \end{cases} .$$

Free parameters for each variable of the feature factor are reduced by regarding them as independent. It is considered that they are interdependent in the performer's individuality; however, determining the shape they take is difficult, and interpretation problems also exist.

The combination of the contextual information derived from $\hat{s}_{n-1}$, $\hat{s}_n$, $\hat{s}_{n+1}$ is associated with the model, based on the assumption that the local context around the direction contributes to the rendering of feature quantities. For the direction about note, various types of information derivable from the score are already under validation as contextual factors [12]. They are primarily in respect of the harmony, which can be regarded as a series consisting of multiple voices and accompaniment, and the main and sub-melodies. According to the orientation of stems of the notes and positional relationship of the chord, each voice part and can be determined automatically and uniquely. Therefore, d_{n-1} and d_{n+1} for d_n are defined with consideration of the structure of the voices and the chords. In the case of the beat, on the other hand, the quantity of information written in a range of one beat to become the observation unit of d_n constantly changes in the score. For models of each beat, directions about rhythm are associated as quantized patterns of keying for each voice and their density, in addition to the global information about the composition.

Refinement of the model with a variety of contextual information is required in order to obtain a context-dependent model that can uniquely describe the rendering process of any case. However, existing performances and the cases

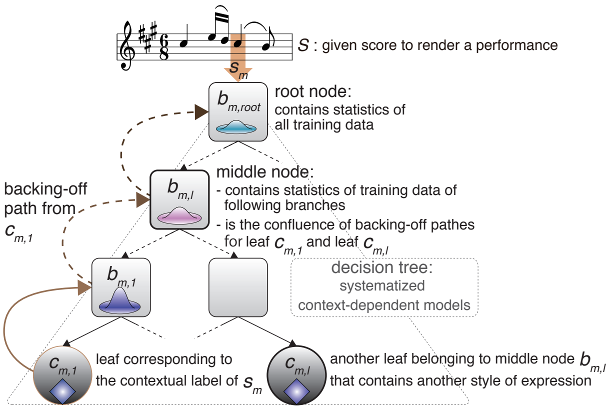


Figure 3: Decision tree backing-off concept.

derivable from them are limited. This means that acquiring models that are able to correlate all the contextual information is effectively impossible. A solution that systematizes the sharing rules is desirable to use as an alternative to any of the finite context-dependent models even for unseen contextual information.

By classifying all context-dependent models using tree-based clustering [14], a decision-tree can be constructed (Figure 2). The structure of the tree elucidates the method by which the case can be rendered with some kind of trend in the performance by the combination of contextual information. Classified context-dependent models for each case are individually arranged at the leaf node of the terminal, and questions about the contextual information become classification criteria and are stored in each intermediate node. It is possible to reach the leaf node of the case with the most similar feature quantities by tracing the intermediate nodes of the tree structure according to each question from the root node. We believe this method effectively identifies known cases with appropriate expression for contextual information of the unknown composition.

2.2 Selecting cases for rendering performance

Owing to the dependence of the optimization criteria of the tree structure on the tendencies of feature quantities and the definition of contextual information, extreme difficulty involved in acquiring the optimum tree structure to render the performance of unseen composition is an issue in the proposal. This means that the corresponding leaf node that is identified by descending the structure with reference to the contextual information is not necessarily the optimum for the performance to render. From examples of analyses obtained in our prior study [12], there is a relatively high versatility that can be commonly explained in the tendency of nodes located near the root of the tree structure. On the other hand, it can be said that nodes near each leaf are specialized in their particular trends. The target of the search for an optimum case should be a subset around the corresponding leaf, and that subset can be augmented by decision-tree backing-off [15]. Candidate cases for the search are gradually augmented from the leaf node $c_{m,1}$, which corresponds to the contextual information of the mth direction s_m of S (Figure 3).

The sequence V is assumed as optimum to render the performance of S. v_m is selected from the candidate cases $C_m = (c_{m,1}, \ldots, c_{m,l}, \ldots, c_{m,L_m})$ that are augmented by the backing-off. If it is assumed that these selections are allowed for each of s_m, dynamic programming [16] may be applied for this search according to the principle of optimality (Figure 4).

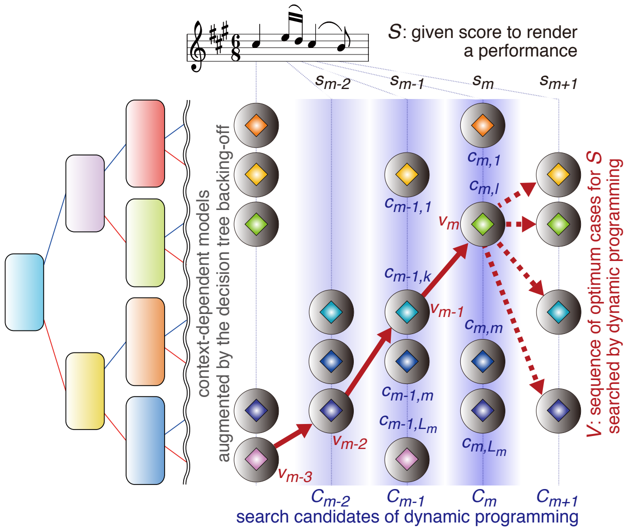


Figure 4: Dynamic programming to select cases that constitute the performance sequence V.

The likelihood based on the feature quantities $\boldsymbol{d}_{m,l}$ that are found in $c_{m,l}$ and the statistics of the middle node $b_{m,l}$ are used to evaluate the suitability of selecting a case $c_{m,l}$ for s_m. First, selection of a case $c_{m,l}$ for s_m is evaluated by $h_1(c_{1,l})$. Next, selection of a pair of cases $(c_{1,k}, c_{2,l})$ for (s_1, s_2) is evaluated by $h_2(c_{1,k}, c_{2,l})$. This process continues until final evaluation by $h_{M-1,M}$. The formulas used to obtain these evaluation values are shown below:

$$h_1\left(c_{1,l}\right) = P\left(\boldsymbol{d}^F_{1,l}|\boldsymbol{\mu}^F_{1,l}, \boldsymbol{\sigma}^F_{1,l}\right) = \prod_{f\in F} P\left(d^f_{1,l}|\mu^f_{1,l}, \sigma^f_{1,l}\right), \quad (4)$$

$$h_m\left(c_{m-1,k}, c_{m,l}\right) = P\left(\boldsymbol{d}^F_{m,l}|\boldsymbol{\mu}^F_{m,l}, \boldsymbol{\sigma}^F_{m,l}\right) P\left(\boldsymbol{\Delta d}^F_{m,l}|\boldsymbol{\mu}^{\Delta F}_{m,l}, \boldsymbol{\sigma}^{\Delta F}_{m,l}\right) \quad (2 \le m \le M). \quad (5)$$

$$\begin{cases} F = (\mathrm{AT}, \mathrm{GR}, \mathrm{DR}), & \Delta F = (\Delta\mathrm{AT}, \Delta\mathrm{GR}, \Delta\mathrm{DR}), \quad \text{for note}, \\ F = \mathrm{BR} & \Delta F = \Delta\mathrm{BR}, \quad \text{for beat}. \end{cases}$$

$\boldsymbol{\Delta d}^F_{m,l} = \boldsymbol{d}^F_{m,l} - \boldsymbol{d}^F_{m-1,k}$ are the differential quantities of each F obtained by assuming the selection of $(c_{m-1,k}, c_{m,l})$ for $(s_{m-1,k}, s_{m,l})$. The search for optimum cases can be viewed as a problem of maximizing evaluation values for each direction of S in the objective function described below:

$$J = h_1\left(c_{1,l}\right) + h_2\left(c_{1,k}, c_{2,l}\right) + \ldots + h_M\left(c_{M-1,k}, c_{M,l}\right) \to \max. \quad (6)$$

All cases included in the tree structure can be candidates for the search, since backing-off reaches the root node finally. However, a search targeting all cases is not always necessary because the possibility that one of the cases in a position significantly distant from the correspondent leaf node in the tree structure is selected as the optimum is unlikely. Therefore, more efficient search is also considered in terms of computational cost by controlling the scale of any augmentation of candidate cases. Index value $\theta_{m,l}$ (shown below) is used to determine continuation or termination of the backing-off, and is determined by the threshold defined in advance:

$$\theta_{m,l} = \left(b^{\max}_m - b^{\min}_m\right)^{-1} \left\{P\left(\boldsymbol{d}_{m,1}|\boldsymbol{\mu}_{m,l}, \boldsymbol{\sigma}_{m,l}\right) - b^{\min}_m\right\} \quad (0 \le \theta_{m,l} \le 1). \quad (7)$$

$b^{\max}_m$ and $b^{\min}_m$ are the maximum and minimum values among the likelihoods obtained for each intermediate node and correspondent leaf node $c_{m,1}$. Augmentation of candidates is restricted only to cases that are very close to the correspondent leaf node if the threshold is close to $\theta_{m,l} = 1$.

Dataset	G(1)	P(1)	P(2)
Performer	G. Gould	M. J. Pires	
Data to train context-dependent models	W. A. Mozart's Piano Sonata, the second and third movements of K. 279 and the first movement of K. 310.		W. A. Mozarts Piano Sonata, the second and third movements of K. 310.
Amount of data	$N^{\mathrm{note}} = 2305$, $N^{\mathrm{beat}} = 396$.	$N^{\mathrm{note}} = 2292$, $N^{\mathrm{beat}} = 396$.	$N^{\mathrm{note}} = 2475$, $N^{\mathrm{beat}} = 504$.

Table 1: Datasets used for the training of context-dependent models.

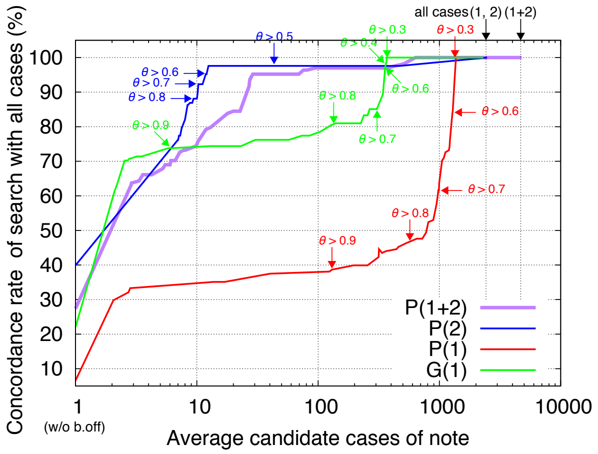


Figure 5: Concordance rate of selected cases for note.

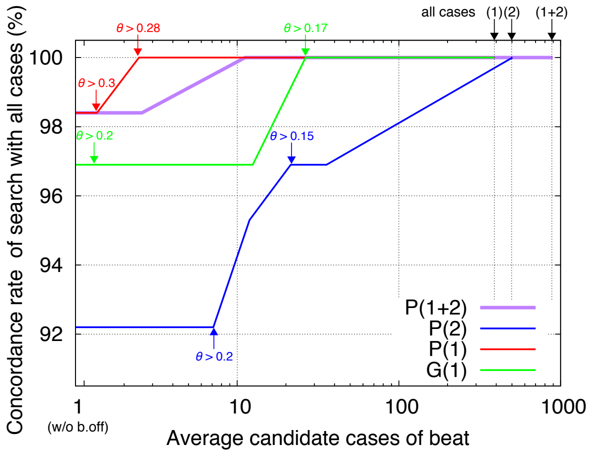


Figure 6: Concordance rate of selected cases for beat.

3. EVALUATION OF THE SYSTEM

We implemented a system for performance rendering based on the proposed method and evaluated the rendered performances from plural terms. Datasets used for the training of context-dependent models were obtained from a database [1] created by musical dictation of the waveforms of a number of virtuosi's piano solo performances on specific MIDI sound generators. The database contains such data related to note and beat converted to the format described above. Directions of the scores were converted to the data that are associable with performance expression by using MusicXML.

3.1 Objective evaluation

In order to verify the effectiveness of the decision tree backing-off method, a number of performances with cases selected by differently scaled backing-off were rendered. The scale of the cases to augment as candidates to search was controlled by the criteria shown in Equation (7). The score used to render the performance here is unified to "W. A. Mozart's Piano Sonata, the first movement of K. 279, (treble voice part)," which is unseen in all the data in the datasets displayed in the Table 1 that are used to train context-dependent models.

For reference assuming a search for all cases of the training data, the matching rates of the cases at the conditions of varied search ranges were examined. The results for note are shown in Figure 5, while those for beat are shown in Figure 6. These figures show that the results of selection in any dataset were exactly matched with the results of "search for all cases," even when the candidates being searched for were limited to only cases from 20% to 50% of all those that are close to $c_{m,1}$ in the decision tree. It can be seen that the cases that are actually effective for any direction are few; thus, effective selection of the case with the optimum expression for such direction from among them should be regarded as important. Decision tree backing-off is a method that allows optimized search of such cases by reducing the number of candidates that need to be examined to find the optimum case for rendering the performance of the unseen direction.

[1] CrestMusePEDB ver. 2, `http://www.crestmuse.jp/pedb`

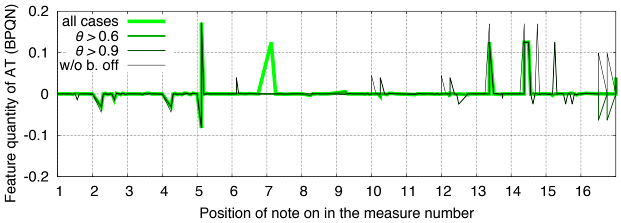
Figure 7 shows the trajectories of the feature quantities for each factor rendered by G(1). In general, similar trends are obtained in terms of each search range. However, for the w/o backing-off condition, there are many cases that fluctuate in the direction opposite to the other conditions and have variations that appear to ignore the trend of all the others. This is a comparison without the correct sequence; however, in general, it is unlikely that such significant local variations without continuity engender naturalness in the performance. The efficacy of introducing decision tree backing-off can also be confirmed from the fact that these strange variations are fixed even in a relatively small augmentation of the search range such as $\theta^{\mathrm{note}}_{m,l} > 0.9$ in note, and $\theta^{\mathrm{beat}}_{m,l} > 0.2$ in beat.

The trajectories of the feature quantities for each factor rendered by P(1), P(2), and P(1+2) are shown in Figure 8. Between these results, the dependence on the combination of the composition and its performer, which is used as the data for training of context-dependent models, is also clear. It can be seen that the tree structure of the models can capture the characteristics of the rendering process of the performance in such combinations. To obtain desirable results for the rendering of unknown compositions, consideration of not only the combination of compositions to train models, but also the difference in characteristics depending on the performer is required. However, clear generalization of the combination and the appropriate amount of training data is difficult to obtain solely from the combination of composition and performer available here. Validation using a context-dependent model separately trained by the combination of cases obtained from a variety of performances is needed.

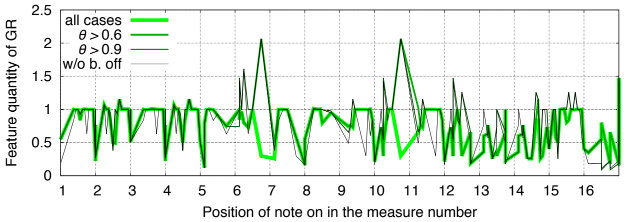
3.2 Subjective evaluation

In order to verify the musical aspects of the performances rendered by the system, they should be evaluated by human listeners. For this evaluation, three performers' models were trained with the data described below:

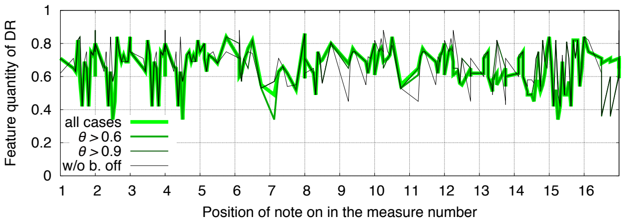
C-A: F. Chopin's Etude Nos. 3, 4, 23; Mazurka No. 5; Nocturne Nos. 2, 10; Prelude Nos. 7, 20; and Waltz Nos. 1, 3, 9, 10, performed by V. Ashkenazy. $N^{\mathrm{note}} = 12092$, $N^{\mathrm{beat}} = 2566$.



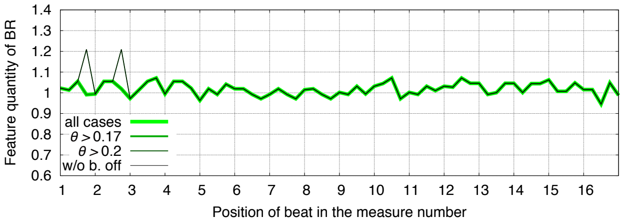
(a) Attack Timing in G(1).



(b) Gatetime Ratio in G(1).

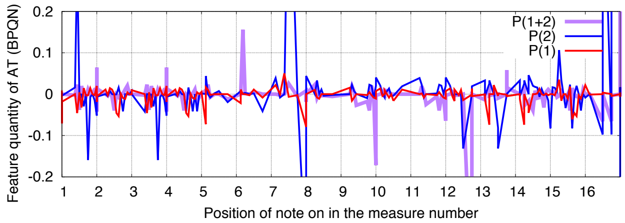


(c) Dynamics Ratio in G(1).

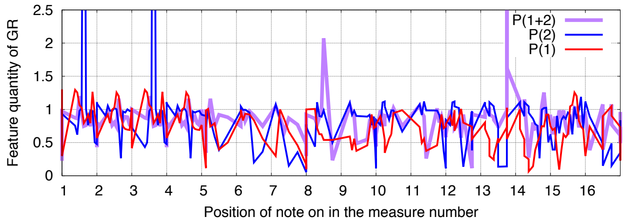


(d) Local BPM Ratio in G(1).

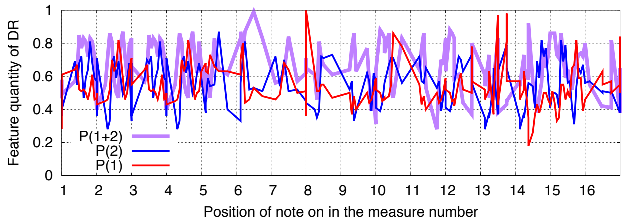
Figure 7: Feature quantities in performances rendered by G(1), for each search range of cases.



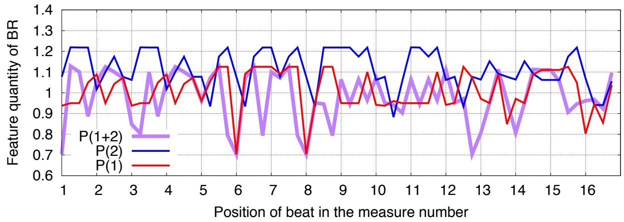
(a) Attack Timing in P(1), P(2), and P(1+2).



(b) Gatetime Ratio in P(1), P(2), and P(1+2).



(c) Dynamics Ratio in P(1), P(2), and P(1+2).



(d) Local BPM Ratio in P(1), P(2), and P(1+2).

Figure 8: Feature quantities in performances rendered by P(1), P(2), and P(1+2), search with all cases.

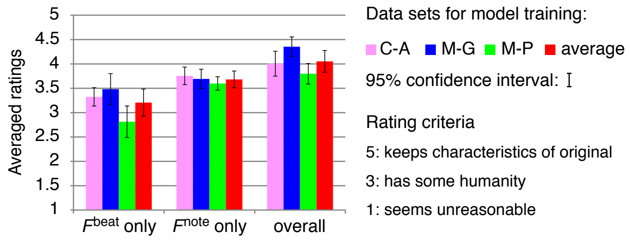
M-G: W. A. Mozart's Piano Sonata, all movements of K. 279 and the first movement of K. 310, performed by G. Gould. $N^{\text{note}} = 3112$, $N^{\text{beat}} = 537$.

M-P: W. A. Mozart's Piano Sonata, all movements of K. 279, K. 310, and K. 545 and the second and third movements of K. 331, performed by M. J. Pires. $N^{\text{note}} = 13703$, $N^{\text{beat}} = 2613$.

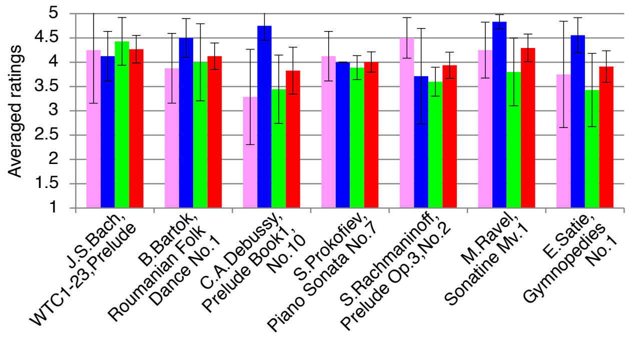
Seven compositions that were not included in the training data and have irrelevant musicality were used for rendering. Twenty participants who were chosen without regard to any professional experience playing musical instruments, evaluated them in five phases. The results obtained for the entire evaluation and metrics used are shown in Figure 9(a). The results obtained by transferring only feature quantity on notes or beats are also shown for reference. Figure 9(b) shows the results evaluated for each composition.

In general, the results obtained are good, as evidenced by the overall evaluation shown in Figure 9(a) having an approximate value of four. These values are generally higher than those obtained for the condition in which only feature quantities related to note are transferred, but the trend is also seen to follow the results for the condition in which only the feature quantity related to beat is transferred. In the M-P model, there is a large bias relative to the contribution to the quality of the performance between each limited transcription condition. It is not necessary for their contribution to the performance to always be equal, but the tree structure of the context-dependent model of beat may not perform as well as that for notes as regards optimum for unknown compositions.

In Figure 9(b), more than half of the compositions for M-G have ratings above four. Simply using a lot of cases to construct the tree structure should not be done lightly because extension of similar cases as candidates that only result in marginal difference in the selection of a case is not desirable for search efficiency. The absolute amount of training data used in M-G is less than that of M-P, but the performances of M-G have a tendency of expression that is able to efficiently capture and transcribe their characteristics. Figure 9(b) also shows large differences in the ratings depending on the compositions in C-A. Combinations of contextual information suitable for the description of the control of expression are different in some cases, since the tendency of expression is also different from the difference in characteristics of the composition even in the case of the same performer. Constructing the tree structure by mixing a large number of such cases is unlikely to be expedient for performance rendering of a particular composition. A simple comparison is difficult because of the difference in compositions and performer, but the combination of Classical music used in M-G and M-P is able to render performances with more stable quality than the combination of Romantic music used in M-G and M-P even for compositions with irrelevant musicality.



(a) Averaged ratings for each (or both) model type.



(b) Averaged ratings for each composition.

Figure 9: Subjective evaluation scores.

4. CONCLUSIONS

This paper proposed an autonomous system for automatic performance rendering with high reproducibility of the characteristics of the performer. It uses stochastic models that associate tendencies of expression in the existing performance and their direction notated in the given score. The structure of automatically systematized models enables efficient search for combinations of cases that are optimized for rendering performances.

Objective evaluations conducted indicate that the decision tree backing-off algorithm enabled efficient search of optimum case series for rendering. The subjective evaluation conducted showed that the system is able to render performances stably even for compositions with unconventional style. Consequently, performances rendered by the proposed system won first prize in the autonomous section of a performance rendering contest for computer systems [17]. The quality of this system was also validated via a large-scale subjective evaluation with eighty participants and piano performance experts. The performances rendered on that occasion are available on the web site that summarizes the results [2]. In addition, more samples rendered in a variety of other compositions are available on our web site [3].

Acknowledgments

This research was supported in part by JSPS KAKENHI (Grant-in-Aid for Scientific Research) Grant Number 26730182, and the Telecommunications Advancement Foundation (TAF).

5. REFERENCES

[1] G. De Poli, "Methodologies for expressiveness modeling of and for music performance," *Journal of New Music Research*, vol. 33, no. 3, pp. 189–202, 2004.

[2] G. Widmer and W. Goebl, "Computational models of expressive music performance: The state of the art," *Journal of New Music Research*, vol. 33, no. 3, pp. 203–216, 2004.

[3] C. Palmer, "Music performance," *Annual review of psychology*, vol. 48, no. 1, pp. 115–138, 1997.

[4] A. Gabrielsson, "Music performance research at the millennium," *Psychology of music*, vol. 31, no. 3, pp. 221–272, 2003.

[5] G. Grindlay and D. Helmbold, "Modeling, analyzing and synthesizing expressive piano performance with graphical models," *Machine Learning Journal*, vol. 65, no. 2-3, pp. 361–387, 2006.

[6] S. Flossmann, M. Grachten, and G. Widmer, "Expressive performance rendering: Introducing performance context," in *Proceedings of Sound and Music Computing (SMC) Conference*, Porto, Portugal, July 2009, pp. 155–160.

[7] T. H. Kim, S. Fukayama, T. Nishimoto, and S. Sagayama, "Performance rendering for polyphonic piano music with a combination of probabilistic models for melody and harmony," in *Proceedings of Sound and Music Computing (SMC) Conference*, Barcelona, Spain, July 2010.

[8] T. Suzuki, T. Tokunaga, and H. Tanaka, "A case based approach to the generation of musical expression," in *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, July 1999, pp. 642–648.

[9] K. Hirata and R. Hiraga, "Ha-Hi-Hun: Performance rendering system of high controllability," in *Proceedings of the ICAD 2002 Rencon Workshop*, Kyoto, Japan, July 2002, pp. 40–46.

[10] A. Tobudic and G. Widmer, "Relational IBL in classical music," *Machine Learning Journal*, vol. 64, no. 1-3, pp. 5–24, September 2006.

[11] J. A. Sloboda, "The acquisition of musical performance expertise: Deconstructing the "talent" account of individual differences in musical expressivity," in *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games.*, K. A. Ericsson, Ed. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., December 1996, ch. 4, pp. 107–126.

[12] K. Okumura, S. Sako, and T. Kitamura, "Stochastic modeling of a musical performance with expressive representations from the musical score," in *Proceedings of the 12th International Society for Music Information Retrieval conference*, Miami, FL, USA, October 2011, pp. 531–536.

[13] C. E. Seashore, *Psychology of Music*. Courier Dover Publications, 1938.

[14] J. J. Odell, "The use of context in large vocabulary speech recognition," Ph.D. dissertation, Cambridge University, 1995.

[15] S. Kataoka, N. Mizutani, K. Tokuda, and T. Kitamura, "Decision-tree backing-off in HMM-based speech synthesis," in *Proceedings of the 8th International Conference on Spoken Language Processing*, Jeju Island, Korea, October 2004, pp. 1205–1208.

[16] R. E. Bellman, *Dynamic Programming*. Princeton University Press, 1957.

[17] K. Okumura, S. Sako, and T. Kitamura. A stochastic model of artistic deviation and its musical score for the elucidation of performance expression. Rencon Working Group. [Online]. Available: http://smac2013.renconmusic.org/participants

[2] Results | Rencon 2013, `http://smac2013.renconmusic.org/results`

[3] Laminae Articulates Musicians' Intention 'N Artistic Expression, `http://www.mmsp.nitech.ac.jp/~k09/laminae`