A Music Search System for Expressive Music Performance Learning

Tomoya Mikami[1,*] and Kosuke Takano[2]

[1] Graduate School of Information and Computer Sciences,
Kanagawa Institute of Technology, Japan
s1385009@cce.kanagawa-it.ac.jp
[2] Department of Information and Computer Sciences,
Kanagawa Institute of Technology, Japan
takano@ic.kanagawa-it.ac.jp

Abstract. In this paper, we present a music search system that focuses on performance style to cultivate a pupil's expressive performance of music. The system allows pupils to learn the performance style to be mastered by obtaining both model and non-model content. By browsing non-model content that is similar to the quality of a pupil's performance, the pupil can quickly identify his/her areas that require improvement. In addition, the pupil can improve his/her performance skill by repeatedly imitating the models. We evaluate the capabilities of our music search system regarding the extraction of performance style from a classical music source and the precision of the music search results for performance style.

Keywords: music retrieval, performance expression, musical performance learning, MIDI.

1 Introduction

In elementary school music education, it is important to foster not only a basic ability of musical activities but also love, sensibility, and sentiment for music by appreciation and expression. For example, when teaching students to play instruments, teachers should encourage pupils to develop the ability to perform based on the musical elements and mood by listening to models as well as understanding the musical score.

In this study, we present a music search system that focuses on performance style to cultivate a pupil's expressive music performance [1, 2]. Our system consists of two main components: performance style extraction using a pre-defined set of rules and a music search that is based on the performance style. The first component extracts the performance style from MIDI data by analyzing the MIDI sequence of a performance. The music search component then retrieves music data based on the performance style extracted from the user's real-time performance with a MIDI instrument.

* Corresponding author.

S. Yamamoto (Ed.): HIMI 2014, Part II, LNCS 8522, pp. 80–89, 2014.

Our system provides two types of music content to pupils for their music performance learning. One is model music that matches the pupil's goal of expressive music performance, and the other is non-model music that is similar to the pupil's current performance. Using our system, pupils can learn the performance style to be mastered by repeatedly listening to the models and imitating them. Meanwhile, by browsing the non-model music, the pupil can quickly identify his/her areas that require improvement. Thus, our system fosters a rich expression of the pupil's music performance through the learning process using both the models and the non-models that are suggested by a search results.

In this study, we evaluate the capability of our system in extracting the performance style from a classical music source and the precision of the music search results of performance style.

2 Learning Scenario

Figure 1 shows an example of a learning scenario using our music search system. To suggest music content (audio and movie) to a student, the system returns both model music and non-model music content based on the student's performance style, which

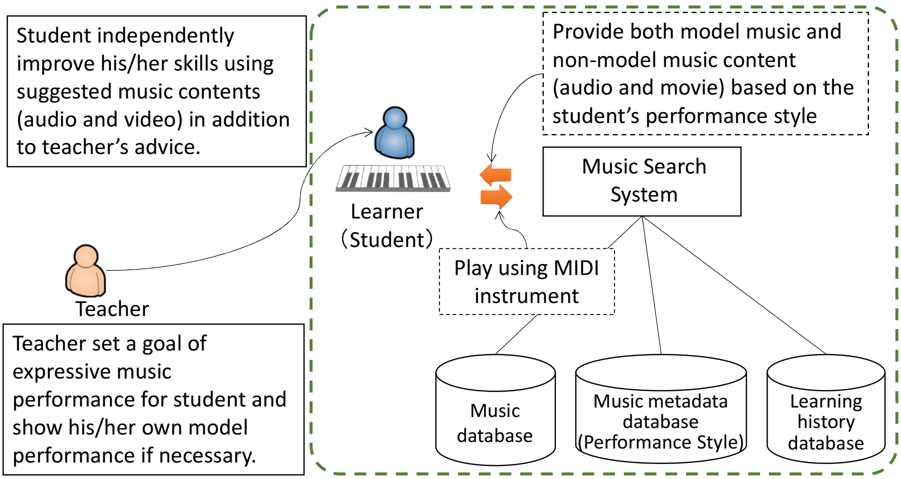

Fig. 1. Example of a learning scenario

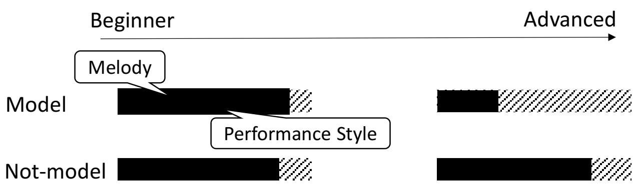

Fig. 2. Parameters of the music search

is extracted from an actual performance using a MIDI instrument. When a teacher indicates negative points about the student's performance, the teacher can make the student logically understand the point. In such a case, the teacher might suggest non-model music that is similar to the current quality of the student's performance so the student can quickly understand the areas that require improvement. Thus, our music search system allows the student to independently improve his/her skills using suggested music content in addition to the teacher's advice.

We incorporate two attributes in the music search: melody and performance style. If the melody attribute is determined to be important, our system retrieves similar music content based on the melody. In contrast, if a student is interested in the performance style, the system suggests similar music content based on the performance style. For example, when a beginner uses the system to browse relevant music content, he/she might be interested in obtaining model and non-model content based on the melody because it is easier for a beginner to improve his/her performance in the early stage of learning by studying music with the same melody. An advanced learner might put more weight on the performance style because he wants to compare his performance to several types of music (Figure 2).

3 Related Work

Content-based music retrieval, such as querying by humming or singing [3, 4, 5, 6, 7, 8], is one of the major approaches for retrieving music content based on acoustic data. Kuo et al. propose four types of query specifications for melody style queries and a melody style mining algorithm to obtain the melody style classification rules for finding music that a user has not listened to [4]. Dynamic Time Warping (DTW) has been used to calculate the similarity of a sequence of pitches in a time series [6]. A hybrid approach using both acoustic and textual features is proposed in [7]. Yang et al. propose a content-based music retrieval algorithm that can be decomposed and parallelized in a peer-to-peer environment. Another approach for music information retrieval is impression or expression-based searches [9], in which impression or emotional terms are extracted from music content, and a user can search for music using terms according to the user's feelings.

Because human feelings, such as impressions and emotion to music, are significant for various music activities, many computer-assisted music systems have been proposed to model a user's emotions, visualize impressions, create expressive performances, and other uses [10, 11, 12]. In [10], machine learning techniques are used to evaluate how expressive performances represented by selected features are clustered in a low-dimensional space. Neocleous et al. propose an emotional modeling technique for music performances that is based on the analysis of how a professional musician represents emotions such as happiness, sadness, anger and fear in violin performances [12].

In addition, many researchers have discussed the benefit of computer-assisted music education in various learning contexts [13, 14, 15, 16]. Through the interactive learning process of experiencing music impressions, students can increase their understanding of important factors that affect the mood of music, such as tonality, rhythm, tempo, pitch, melody, and harmony [13]. Morijiri discusses the feedback effect that listening to recordings of their own performances has on piano performers and states that performers can improve by listening to their recordings. Ng [15] proposes an interactive multimodal feedback system in which the playing gesture is visualized using 3D motion capture technology. Furthermore, Ogura discusses the benefit of providing a model piano performance for beginners, especially in terms of rhythm, taste of music, tempo, and fingering [16].

4 Proposed System

Figure 3 shows an overview of the proposed music search system. A pupil first inputs the performance information as a search query by playing a MIDI instrument. The music search system then provides both model music content and non-model music content.

In the following sections, we describe the main components of the system, which include a performance style extractor and a music search engine that focuses on the performance style.

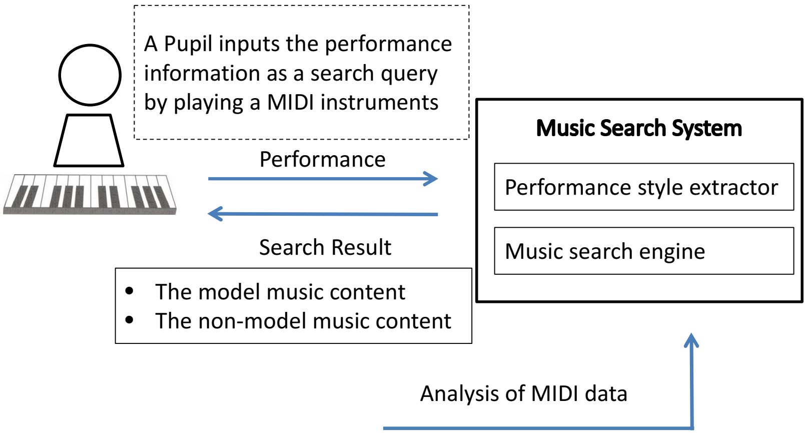

Fig. 3. Overview of our system

4.1 Performance Style Extraction

To retrieve music content to improve a pupil's expressive music performance, we extract the performance style from MIDI data.

Several music symbol types (tempo, dynamics, and style) are used to evaluate performance in musical expression. In particular, style is used to instruct emotional expression for music performance. We choose 38 performance styles from the music expression marks; several styles are shown in Table 1.

The performance styles are extracted based on a pre-defined set of rules that analyzes the values of expressive MIDI controls, such as velocity, modulation, and sustain. Velocity (0 to 127) is a MIDI control of loudness; higher values indicate louder sounds. Modulation (0 to 127) provides a sound fluctuation effect like *vibrato*. In addition, sustain (on or off) provides the sound duration. We design a rule set for performance style extraction using these MIDI controls. Examples of these rules are shown in Table 2. The extracted music performance style is used for the music search process.

Table 1. Examples of performance styles

ID	Style	Meaning
f_1	grandioso	with grandeur
f_2	doloroso	sorrowfully
f_3	cantabile	in a singing style
f_4	furioso	furious
f_5	espressivo	expressively
f_6	schwach	composed
f_7	placido	pacific
f_8	staccato	short and detached
f_9	altisonante	harmonically

Table 2. Examples of performance extraction rules

Performance style	**MIDI control**	**Rule**
Grand	Velocity	100 <= average
	Modulation	Not considered
	Sustain	Not considered
Estito	Velocity	(1) 100 <= average and 5 < variance < 10 (2) 60 <= average <= 100 and variance <= 5
	Modulation	Not considered
	Sustain	Not considered
Altisonante	Velocity	A few notes have average differences that are greater than 10
	Modulation	More than once
	Sustain	No considered
Piangendo	Velocity	A few notes have average differences that are greater than 20
	Modulation	More than once
	Sustain	Not considered

4.2 Music Search Based on Performance Style

In the music search process, the system calculates the similarity between input MIDI performance data $MIDI_q$ and each MIDI data file $MIDI_i$ in the music database. The similarity calculation is performed as follows:

Step 1: The MIDI data are vectorized to $\mathbf{style}_q$ and $\mathbf{style}_i$, respectively, by applying the performance rule shown in Table 2. The feature vectors $\mathbf{style}_q$ and $\mathbf{style}_i$ are 38 dimensional vectors that are featured using the 38 performance styles f_x. Examples of performance styles are shown in Table 1.

Step 2: The similarity of the performance style between $\mathbf{style}_q$ and $\mathbf{style}_i$ is calculated using a cosine measure.

$$SIM_{style}(\mathbf{style}_q, \mathbf{style}_i) = (\mathbf{style}_q \cdot \mathbf{style}_i) / |\mathbf{style}_q||\mathbf{style}_i| \tag{1}$$

Step 3: The similarity of the melody between $MIDI_q$ and $MIDI_i$ is calculated using a Dynamic Time Warping (DTW) algorithm [6, 17].

$$SIM_{melody}(\mathbf{melody}_q, \mathbf{melody}_i) = DTW(\mathbf{melody}_q, \mathbf{melody}_i) \tag{2}$$

Step 4: Finally, the similarity between $MIDI_q$ and $MIDI_i$ is calculated as the summation of SIM_{style} and SIM_{melody}.

$$SIM(MIDI_q, MIDI_i) = \alpha \cdot SIM_{style}(\mathbf{style}_q, \mathbf{style}_i) + \frac{1-\alpha}{DTW(\mathbf{melody}_q, \mathbf{melody}_i)} \quad (0 \le \alpha \le 1) \tag{3}$$

When we apply equation (3) in the music retrieval process, the parameter α should be set to a proper value according to the learning level and purpose. For example, when a pupil searches for the model music content in the early stages of learning, the search system should emphasize the melody factor because it is easier for the pupil to follow an example. After the learning is repeated, the pupil might want to listen to several pieces with similar performance expressions but different melodies; in this case, the search system should emphasize the performance style factor.

5 Experiment

We evaluated the performance of our system based on (i) the accuracy of the performance style extraction from a MIDI classical music source and (ii) the precision of the music search based on the extracted performance style.

5.1 Experiment 1

In experiment 1, a skilled musician created 75 MIDI files of famous classical music with different performance styles. To evaluate the accuracy of the performance style extraction, we used short music clips (average time: 12 seconds) with the same melodic parts from one classical music song (title: *Amazing Grace*) to create the 75 MIDI files. For the evaluation, the same musician defined the correct performance style of each MIDI file by listening to them. We calculate precision and recall as follows:

Recall = (Number of performance styles extracted by the proposed rule)
/ (Number of correct performance styles estimated by the musician)

Precision = (Number of performance styles)
/ (Number of performance styles extracted by the proposed rule)

Table 3. Examples of performance styles extracted by the pre-defined rules

Music ID	Extracted performance style	Correct performance style
Amazing 24	Expressively	Calm
Amazing 44	With grandeur, expressively, harmonically	With grandeur, expressively
Amazing 63	Calm, extinguished	Calm, extinguished

Table 4. Average accuracy of performance style extraction

Average Recall	Average Precision
0.68	0.49

Table 3 shows examples of the performance style extraction results using the rules defined in Table 2. The proper performance styles are extracted for Amazing Grace 44 and 63. However, for Amazing Grace 24, the performance style "Expressively" is extracted instead of "Calm". The reason for this result is that the person who defined the correct performance style could not perceive the expressivity by ear because Amazing Grace 24 is too calm, although it is rich in variation.

Table 4 shows the average accuracy of the performance style extraction. The average recall is high, but the average precision is slightly lower; this is because our system cumulatively extracts the proper performance styles that match the extraction rule as shown in "Amazing Grace 44" in Table 3.

5.2 Experiment 2

In this experiment, we used 60 MIDI files to evaluate the precision of the music search based on the extracted performance style. The 60 MIDI files are short music segments (average time: 9.34 seconds) from three classical music songs (*Amazing Grace*, *Hungarian Dances No.5*, and *Fauré by Pelléas et Mélisande*) that include 20 different performance styles. For the evaluation, we input each MIDI file as a query and ranked the MIDI files based on a similarity score, which was calculated as described in Section 4.2. We set the parameter α in equation (3) to 1, so we can evaluate only the music search based on the performance style. In addition, we define the correct music content of each query MIDI file as follows:

1. The same musician from Experiment 1 defined the correct performance style for each MIDI file.
2. Each MIDI file was vectorized based on the performance styles, and their similarity scores were calculated as described in Section 4.2.
3. We defined the correct MIDI file m_c for a query MIDI file m_q, where the similarity score between m_c and m_q is greater than 0.8.

Tables 5 and 6 show the ranking results for music queries Amazing 13 and Faure 10, respectively. The performance styles of Amazing 13 extracted by our rule are expressively, calm, and short and detached. Table 5 shows that our method can identify the proper music content based on the performance style of Amazing 13 and not by the melodic similarity. Similarly, Table 6 shows that our method can obtain proper music content for the query music file Faure 10, whose performance styles are calm, raging, and short and detached.

In addition, Figure 4 shows the average recall precision rate of the search results using all 60 MIDI files as queries. These results confirm that our system can retrieve appropriate music content based on the performance style that a user plays on a MIDI device.

Table 5. Top 5 results (Query music: Amazing 13)

Rank	Music ID	Score	Extracted performance style	Correct performance style
1	Hungarian 16	0.816	Expressively, short and detached	Expressively, calm
2	Hungarian 17	0.707	Raging, short and detached, pacific	Expressively, pacific, short and detached
3	Amazing 1	0.707	With grandeur, short and detached	Expressively, such as tempo, short and detached
4	Amazing 6	0.707	Expressively, short and detached, such as tempo, mournful	Expressively, such as tempo, short and detached
5	Faure 14	0.632	Expressively, such as tempo, short and detached	Short and detached, quick and lively

Table 6. Top 5 results (Query music: Faure 10)

Rank	Music ID	Score	Extracted performance style	Correct performance style
1	Amazing 16	0.816	Calm, expressively, mournful	Expressively, such as tempo, with grandeur, short and detached
2	Faure 2	0.730	Calm, expressively	Mournful, heavily, spread
3	Faure 9	0.730	Calm, raging, short and detached	Mournful, expressively, spread
4	Amazing 17	0.730	Calm, expressively, spread	Extinguished, such as tempo, with grandeur, short and detached
5	Faure 3	0.707	Calm, expressively	Expressively, spread

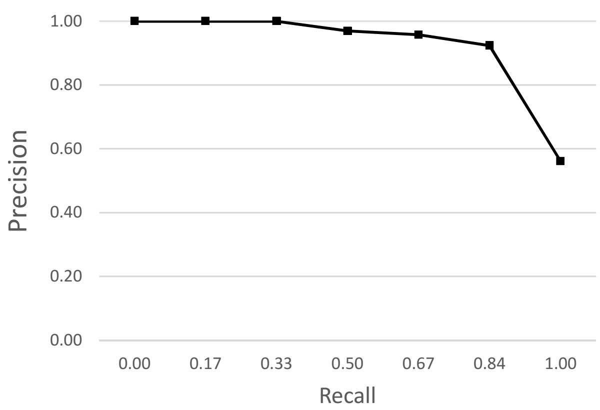

Fig. 4. Average recall precision rate of the search results

6 Conclusion

We presented a music search system that focuses on performance style, in which the performance styles are extracted based on a pre-defined set of rules, and the music contents are retrieved based on the performance style. Our music search system allows a pupil to obtain both model and non-model music content by inputting a performance as a search query by playing a MIDI instrument. Experiments were used to evaluate the accuracy of our system for performance style extraction and music search.

In the future, we will perform experiments using our system at an elementary school. We will validate the effectiveness of the proposed music system and improve the system based on feedback from teachers and pupils.

References

1. Chika, O., Kazushi, N., Yohei, M., Takashi, S.: A Facilitating System for Composing MIDI Sequence Data by Separate Input of Expressive Elements and Pitch Data. IPSJ Journal 44(7), 1778–1790 (2003) (in Japanese)
2. Suzuki, T., Tokunaga, T., Tanaka, H.: A Case-based Approach to the Generation of Musical Expression. IPSJ Journal 44(7), 1778–1790 (2003) (in Japanese)
3. Ghias, A., Logan, J., Chamberlin, D., et al.: Query By Humming – Musical Information Retrieval in an Audio Database. Proceedings of ACM Multimedia 95 (1995)
4. Kuo, F.-F., Shan, M.-K.: Looking for new, not known music only. In: IEEE-CS Joint Conference on Digital Libraries, pp. 243–251 (2004)
5. Zhu, Y., Kankanhalli, M.: Music scale modeling for melody matching. In: Proceedings of the Eleventh ACM International Conference on Multimedia (MULTIMEDIA 2003), pp. 359–362 (2003)
6. Jang, J.-S.R., Lee, H.-R.: Hierarchical Filtering Method for Content-based Music Retrieval via Acoustic Input. In: Proceedings of the Ninth ACM International Conference on Multimedia (MULTIMEDIA 2001), pp. 401–410 (2001)
7. Cui, B., Liu, L., Pu, C., Shen, J., Tan, K.-L.: QueST: querying music databases by acoustic and textual features. In: Proceedings of the 15th International Conference on Multimedia (MULTIMEDIA 2007), pp. 1055–1064 (2007)
8. Yang, C.: Peer-to-peer architecture for content-based music retrieval on acoustic data. In: Proceedings of the 12th International Conference on World Wide Web (WWW 2003), pp. 376–383 (2003)
9. Kitagawa, T., Kiyoki, Y.: Fundamental Framework for Media Data Retrieval Systems Using Media-lexico Transformation Operator. In: The Case of Musical MIDI Data, Information Modeling and Knowledge Bases, vol. 12, pp. 316–326. IOS Press (2001)
10. Mion, L., De Poli, G., Rapana, E.: Perceptual organization of affective and sensorial expressive intentions in music performance. ACM Transactions on Applied Perception 14, 1–21 (2010)
11. Yang, Y.-H., Homer, H.: Machine Recognition of Music Emotion. ACM Transactions on Intelligent Systems and Technology, Article No. 40 (2012)
12. Neocleous, A., Ramirez, R., Perez, A., Maestre, E.: Modeling emotions in violin audio recordings. In: Proceedings of 3rd International Workshop on Machine Learning and Music, vol. 147, pp. 17–20 (2010)
13. Kirke, A., Miranda, E.R.: Survey of Computer Systems for Expressive Music Performance. Survey of Computer Systems for Expressive Music Performance 42(1) (2009)
14. Morijiri, Y.: The effect of self-evaluation on piano performers: using feedback by listening to a recording after performance. Journal of the Graduate School of Humanities and Sciences 12, 111–119 (2009) (in Japanese)
15. Ng, K.: Interactive feedbacks with visualisation and sonification for technology-enhanced learning for music performance. In: Proceedings of the 26th Annual ACM International Conference on Design of Communication, pp. 281–282 (2008)
16. Ryuichiro, O.: A Trial to Utilize a Performance Model for ML Learning: Give the Singing and Playing Video of the Song of the Child to the Learner. Annual Report of the Faculty of Education, Bunkyo University 46, 77–84 (2013) (in Japanese)
17. Myers, C.S., Rabiner, L.R.: A comparative study of several dynamic time-warping algorithms for connected word recognition. The Bell System Technical Journal 60(7), 1389–1409 (1981)