be described as emergent, due to their reliance on tools or individuation in others. In addition, other modalities can as these latter two roles show, design inheres within both composers to design infrastructures which stand apart from each other.

Figure 4 Elements of a Socially Constructed Performance System

5. CONCLUSIONS

In the course of this paper, we have presented an ethno- graphic study of performance modalities in BiLE’s Laptop opera Act 2, identifying the stakeholders involved in each modality and considering how multiple modalities manifest themselves as socially constructed performance systems. Our investigation shows that whilst traditional roles of composer and performer are present within BiLE practice, the consensus driven nature of the group produces a dynamic set of orientations, which do not strictly define players’ roles. Instead, interaction is a dynamic activity, a social process driven by the group. The boundaries of collaboration are negotiated anew — rather than their explicit formalisation — that acts as the primary driving force in the creation of new work.

6. REFERENCES


A SYSTEM FOR MEMORIZING SONGS BY PRESENTING MUSICAL STRUCTURES BASED ON PHRASE SIMILARITY

Yuma Ito
Kobe University

Yoshinari Takegawa
Future University Hakodate
Tsutomu Terada
Kobe University PRESTO, JST
Masahiko Tsukamoto
Kobe University

ABSTRACT

Players of musical instruments usually memorize musical scores for concerts and live performances. However, memorizing songs requires much effort on the part of the player as they have to play and listen to the song over again. The goal of our study is to construct a system for memorizing musical scores based on the phrase similarity. The proposed system calculates the phrase similarity in the target song, and presents the musical structures and the different points in similar phrases based on the phrase similarity. The player can understand the musical structure immediately, and can memorize the musical score in a short time because of the reduction of duplicated learning for the similar phrases. Our evaluation results confirmed that our method had advantages compared with conventional musical scores.

1. INTRODUCTION

Players on musical instruments usually memorize musical scores for concerts and live performances. It is important for musicians to memorize musical scores. However, for memorizing musical scores, it needs a great effort on players by playing and listening the song over again. In addition, because it is difficult to memorize musical scores correctly, players sometimes play the same phrase multiple times or forget part of a song due to stress when performing in front of an audience.

On the other hand, a song has musical structures, such as motifs. Figure 1 shows an example of musical structure. There are multiple layers from abstracted layer to detailed layer as shown in Figure 1. The learner is composed of multiple musical layers from abstracted to detailed. Additionally, each bracket indicates a phrase belonging to each layer. Most of these musical structures are composed of similar motifs. In addition, information on intersection among different points of two similar phrases. In our research, it is especially important for them to be conscious of the different points of two similar phrases. Thus, we develop a system that presents the learner from memorizing ambiguously by presenting different points among similar phrases on musical scores, thus helping the learner to memorize musical scores correctly.

3. DESIGN

As mentioned in Section 2, it is important to be conscious of the different points in similar phrases. Musical structure is composed of multiple musical layers from abstracted layer to detailed layer as shown in Figure 1. The learner requires the abstracted layers to check brief overview of a song. In contrast, the learner requires the detailed layers to check the melody components of a song, which are rhythm, pitch, and musical techniques. Because the necessary layer is dependent on the ability or the interest of
the learner, the proposed system must be able to switch flexibly among presentation of the different layers.

3.1. Phrase Similarity

In this research, the degree of similarity is calculated in units of phrase. The phrases and their hierarchical structure are generated using the method proposed by Hamanaka et al. [4] on the basis of GTTM proposed by Lerdahl et al. [5]. Additionally, the degree of similarity is calculated based on two types of similarities. One is physical similarity, such as fingering, which is dependent on musical instruments, and the other is musical similarity, such as pitch and duration, which is independent of musical instruments.

3.1.1. Musical similarity

The musical similarity is calculated based on feature values such as pitch and rhythm, which are extracted from each phrase. We employ three feature values; timing (the onset time of each note), pitch (the absolute pitch), and interval of pitch (the difference in pitch between the current note and the one that precedes). Figure 2 shows an example of timing. We define the onset time of each note on the basis of the first note of each phrase. Each rectangle in the figure denotes a phrase. In our system, the learner can change the presented layer in the layers freely. The learner uses this mode to learn the structure of the current layer or to select a base phrase used in the Similar phrases presentation mode.

3.2. Method of Memorization

We explain a proposed method for memorizing scores with Figure 3 and Figure 4. The proposed system has two modes: All phrases presentation mode and Similar phrases presentation mode. The learner uses the All phrases presentation mode to learn the structure of the song or to select a base phrase that is used in the Similar phrases presentation mode. Additionally, users learn the similarities and differences between the selected base phrase and other phrases with the Similar phrases presentation mode.

3.2.1. All phrases presentation mode

Figure 3 shows an example of the All phrases presentation mode. Each rectangle in the figure denotes a phrase. In the figure, the learner can recognize the phrase that is not identical to the previous one. Additionally, the colored circles show the degree of similarity. In the right-hand diagram, the green circle corresponds to the degree of similarity having a value of 1, and the red one corresponds to a value of 0. The rightmost scores are guitar tabs. Details of the example in Figure 4 are given below, and the Roman numerals in the black dots in Figure 4 correspond to the following list.

(i) The phrases surrounded by a solid rectangle are base phrases, whereas similar phrases are surrounded by a dotted rectangle. The numbers next to the rectangles show the degree of similarity. In the right-hand diagram of Figure 4, the base phrase is placed at the top of the list, and other similar phrases are arranged in order from the highest degree of similarity to the lowest.

(ii) The number of identical phrases is indicated with a numerical value such as “X2”, which appears next to the base phrase. In this way, duplicated information is reduced because the learner does not have to re-remember the phrase when it reappears later on in the song.

3.2.2. Similar phrases presentation mode

This mode presents a base phrase selected by the learner in the All phrases presentation mode and the phrases that are similar to it. We propose two types of content presentation as shown in Figure 4. The left-hand diagram shows a general musical score, and the right-hand diagram shows a summary of the specific phrases. The rightmost scores are guitar tabs. Details of the example in Figure 4 are given below, and the Roman numerals in the black dots in Figure 4 correspond to the following list.

(i) When the timing or fingering of a note in a similar phrase is the same as that of the base phrase, they are connected by a dashed line. The learner can understand the similarity of timing or fingering.

(ii) When there are the notes in the similar phrases that are same as those of the base phrase in regard to pitch, timing, and fingering, they are surrounded by circles. The learner can understand the similarity of each feature value.

3.3. Physical similarity

Fingering in playing a musical instrument is an example of physical similarity. Because positional and physical information, such as fingering, is almost mechanical memory, it is important to back this up with conscious memory by presenting the fingering similarity. Additionally, fingering is different for each musical instrument. In other words, physical information such as fingering is dependent on the musical instrument.

Physical similarity is calculated by DTW in the same way as musical similarity. The physical distance between i-th and j-th is defined as \(d_\pi (i, j)\). Strings, frets, and musical performance techniques such as arpeggio and hammer on are an element of the DTW of physical similarity. \(d_\pi (i, j)\) is the sum total of the results of DTW for each element. Two phrases of which the \(d_\pi (i, j)\) is lower than the threshold are defined as physically similar phrases.

4. IMPLEMENTATION AND EVALUATION

We implemented a prototype system for memorizing musical scores. The prototype stores musical score including meta-data such as pitch, timing, and fingering, in XML. It calculates the degree of the similarity among phrases based on the musical score data. We implemented the prototype using Microsoft Visual C# 2008 on Windows XP.

We evaluated the proposed method by comparing with a conventional method.

In the evaluation, the subjects practiced the song until they had memorized it using the two methods: the conventional method, which uses conventional musical scores, and the proposed method, which uses the musical scores created by the prototype. We measured how long subjects took to memorize each trial score. Each test subject memorized four; two songs with the conventional method and two songs with the proposed method. The subjects were permitted to use the electric guitar and listen to the songs during the evaluation as needed. Furthermore, before the subjects worked on the task with the proposed method, we instructed them on how to operate the prototype and how to learn songs with it.

All the songs we used belong to the rock music genre and they are composed of approximately 6 to 9 bars. We call these songs Song A, Song B, Song C and Song D for convenience.

There were four subjects, all of whom were university students in their early 20s who major in engineering, can read musical scores, and play the electric guitar as well. All the subjects had never listened to or practiced the trial songs before.

Table 1 shows the results of the experiment. Each value in Table 1 denotes the minutes that a subject took to memorize a song, which is referred to as memorization time. The gray cells indicate the memorization time applied to the conventional method, and the white cells indicate the memorization time applied to the conventional method. Each subject had different level of ability for reading musical scores and different musical techniques. Each song presented different difficulties. Therefore, we normalize the memorization time based on the memorization time applied to the conventional method. Specifically, we define the ability of each subject as follows:

\[a_s(n) = n \leq 2, 3, 4\]

Note that “n” corresponds to the number assigned to each subject. We determine the memorization time of the conventional method. For example, since Subject 1 memorized Song C and Song D using the conven-
3.1.1. Musical similarity
The musical similarity is calculated based on feature values such as pitch and rhythm, which are extracted from each phrase. We employ three feature values: timing (the onset time of each note), pitch (the absolute pitch), and interval of pitch (the difference in pitch between the current note and the one that precedes).

Figure 2 shows an example of timing. We define the onset time of each note on the basis of the first note of a phrase as timing. For example, the two phrases shown in the figure have the same onset timing in, as shown by the red notes in the figure.

Our system uses DTW (Dynamic Time Warping) [6] to measure the similarity of two phrases. DTW can be used to measure the similarity between two sequences, each of which may be differently stretched or compressed in time. The value of the Euclidean distance $d_{w}(i,j)$ is defined by the following equation. $N$ is the number of phrases in a musical layer.

$$d_{w}(i,j) = w_{p}d_{p}(i,j) + w_{r}d_{r}(i,j) + w_{f}d_{f}(i,j)$$

(i) The phrases surrounded by a solid rectangle are base phrases, whereas, similar phrases are surrounded by a dotted rectangle. The numbers next to the rectangles, show the degree of similarity. In the right-hand diagram of Figure 4, the base phrase is placed at the top of the list, and other similar phrases are arranged in order from the highest degree of similarity to the lowest. The learner can easily understand the similar phrases and their location in the musical score.

(ii) The number of identical phrases is indicated with a numerical value, such as “x2,” which appears next to the base phrase. In this way, duplicated information is reduced because the learner does not have to re-remember the phrase when it reappears later on in the song.

3.2.3. Similar phrases presentation mode
This model presents a base phrase selected by the learner in the All phrases presentation mode and the phrases that are similar to. We propose two types of content presentation as shown in Figure 4. The left-hand diagram shows a general musical score, and the right-hand diagram shows a summary of the similar phrases. The rightmost score is a guitar tab. Details of the example in Figure 4 are given below, and the Roman numerals in the black dots in Figure 4 correspond to the following list.

(iii) When the timing or fingering of a note in a similar phrase is the same as that of the base phrase, they are connected by a dashed line. The learner can understand the similarity of timing or fingering.

(iv) When there are the notes in the similar phrases that are same as those of the base phrase in regard to pitch, timing, and fingering, they are surrounded by circles. The learner can understand the similarity of each feature value.

4. IMPLEMENTATION AND EVALUATION
We implemented a prototype system for memorizing musical scores. The prototype stores musical score including meta-data such as pitch, timing, and fingering, in XML. It calculates the degree of the similarity among phrases based on the musical score data. We implemented the prototype using Microsoft Visual C# 2008 on Windows XP.

We evaluated the proposed method by comparing with a conventional method.

Table 1. Memorization time

<table>
<thead>
<tr>
<th>Song</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub.1</td>
<td>0.67</td>
<td>1.21</td>
<td>0.85</td>
<td>1.69</td>
<td>2.80</td>
</tr>
<tr>
<td>Sub.2</td>
<td>1.56</td>
<td>2.13</td>
<td>1.15</td>
<td>1.43</td>
<td>2.94</td>
</tr>
<tr>
<td>Sub.3</td>
<td>0.89</td>
<td>1.05</td>
<td>0.96</td>
<td>1.30</td>
<td>2.55</td>
</tr>
<tr>
<td>Sub.4</td>
<td>0.86</td>
<td>1.20</td>
<td>1.41</td>
<td>0.71</td>
<td>1.48</td>
</tr>
</tbody>
</table>

We measured how much time each subject took to memorize a song, which is referred to as memorization time. The gray cells indicate the memorization time applied to the proposed method, and the white cells indicate the memorization time applied to the conventional method.

Each subject had different level of ability for reading musical scores and different musical techniques. Each song presented different difficulties. Therefore, we normalized the memorization time based on the proposed method to the conventional method. Specifically, we define the ability of each subject as follows:

$$a_{s}(n = 1, 2, 3, 4)$$

Note that “n” corresponds to the number assigned to each subject. We determine the memorization time of the conventional method. For example, Subject 1 memorized Song C and Song D using the conven-
sional method, we can compare Subject 1 with Subject 2 and Subject 4 who also memorized Song C or Song D using the conventional method. The equation is as follows:

\[
\alpha = \frac{16}{251} = \frac{19}{274}
\]

We formulate the equation for each subject, and define the average number of multiple solutions of \(\alpha_i\) as the final \(\alpha_i\), as shown in the rightmost in the table.

Furthermore, we calculate the primary memorization time in the case in which the subject did not use the proposed method in contrast to the memorization time with the proposed method. The primary memorization time is determined on the basis that the ratio of the primary memory time to the time using the conventional method is equal to the ratio of abilities. For example, if we compare Subject 1 with Subject 2, the primary memorization time \(x\) is determined as 15.6 by following equation.

\[
x = \frac{4.00}{1.03} = 15.6
\]

This value indicates that the primary memorization time of Subject 1 was reduced by 15.6 min to 11 min by using the proposed method.

We determine the ratio of the memorization time of the proposed method to primary memorization time, and we refer to this as the ratio of memorization time. The ratio of memorization time is determined by the following equation.

\[
x = \frac{11}{15.6} = 0.71
\]

This value indicates that the memorization time is reduced by approximately 30%.

Table 2 shows the ratio of memorization time in all the combinations. The three columns on the right-hand side and the bottom three rows of Table 2 show the value of average, standard deviation, and p-value that indicates the statistical significance by t-test for each subject and each song.

The memorization time of Song A and Song C are reduced. The proposed system worked effectively on these songs. The reason is that over half of the phrases are used. The proposed system worked effectively on these songs. The reason is that over half of the phrases are used. The proposed system worked effectively on these songs. The reason is that over half of the phrases are used. The proposed system worked effectively on these songs. The reason is that over half of the phrases are used.