Music Similarity Retrieval Method Considering Music Arrangement

Kenji Kogo
Department of Information Science
Kyoto Institute of Technology
Kyoto, Japan

Kyoji Kawagoe
Dept. of Information and Communication Science
Ritsumeikan University
Shiga, Japan

Teruhisa Hochin
Information and Human Sciences
Kyoto Institute of Technology
Kyoto, Japan
Email: hochin@kit.ac.jp

***Abstract*—This paper proposes a music similarity retrieval method considering the shift of the pitch and the difference of tempos of music pieces. These are not considered in the current various services yet. The proposed method can retrieve music pieces arranged from the original ones as similar ones. The proposed method uses the pitch information and the length of the sound. The pitch and the tempo are normalized in calculating the similarity of music pieces. The effectiveness of the proposed method is experimentally shown.**

I. Introduction

Internet music providing services like iTunes or Hi-Resolution [1] are very popular because many people usually use potable music players today. Many kinds of services are provided. We can find musics by specifying artist names, song titles [2], or lyrics [3][4]. Many scientists have studied about Music Information Retrieval (MIR) [5][6].

On the other hand, there are many arranged musics, e.g., popular music based Classic pitches, and a dance music arranged in jazz music. However, the music retrieval system which can get musics of different genres or different artists from the original one does not exist. For example, a user is listening to an music, and he likes it. Assume it created by arranging an original music. He cannot search the original music as a similar one through traditional music retrieval systems. An arranged music is judged different from the original music. To solve this problem, it is required to acquire and analyze the music information such as lengths and pitches of sound.

This paper proposes a melody search technique that music provided in the arrangement from an original music are obtained as the musics similar to the original one. This paper treats Standard Music Instrumental Digital Interface (MIDI) files (SMFs) as music files. An SMF stores the information of notes including pitches and lengths as codes. The proposed method uses pitches of melody information, and lengths of the sound. In this method, pitches and lengths are handled as time-series data. The similarity between timed-series data is calculated. The Dynamic Time Warping (DTW) method, which is often used in measuring the similarity of time-series data, is used. However, this method cannot consider the difference of the key between an original music piece and an arranged one, and that of the tempo between them. Therefore, the normalization of time-series data is adopted to solve these problems. In addition, this paper proposes a method of identifying portion of an arranged music piece partially matched to the original one. This method is based on a sliding window method.

The remainder of the paper is organized as follows : Section 2 describes related works on the similarities of musics based on melody and audio information. Section 3 proposes a music retrieval method. Evaluation experiments are conducted in Section 4. Finally, Section 5 concludes the paper.

II. Related Research

The studies on conventional musical piece search and melody search are described. First, the musical piece search technique using a musical characteristic and the problems are described. Next, the melody search system used in this study and the problems are described.

A. System of Music Retrieval

Pesek *et al.*[7] suggested a hierarchical structure model of the musical piece frequency as a new model of the musical piece search system. They acquired audio information of musical pieces and analyzed which frequency band appeared in a musical piece. Their technique makes the hierarchical structure consisting of three levels based on the appearance frequency of the frequency. It builds a neural network between hierarchies, and performs Unsupervised Learning. They show possibility of the musical piece presentation by using the technique of the field of such artificial intelligence. However, problems of this study are to pay attention to only appearance frequency of the sonic frequency. It is possible to estimate the mood of the music by analyzing the amplitude increase and decrease of waveforms of all frequencies. This method could, however, estimate the similarities of only instruments used and tones. This paper focuses on pitches and lengths, and pays attention to the similarity of the melody flow. This makes it possible to judge an arranged music piece similar to the original one.

On the other hand, Grosche *et al.*[8] extract quantity of characteristic from music structure and argue that they can measure similarity with a simple distance standard. As it can calculate similarity without using the distance standard like DTW or Longest Common Subsequence (LCS), but using the Euclid distance, calculation cost decreases. However, they do

978-1-5090-0806-3/16/$31.00 copyright 2016 IEEE
ICIS 2016, June 26-29, 2016, Okayama, Japan

not touch it about partial similarity at all. In other words, it may be treated as a noise because they watch only a resemblance degree as the whole even if there was the music that it is partially similar. Therefore, this paper calculates the partial similarity by using a sliding window method. It is shown that a music piece is partially similar to the melody by reflecting this result when the a final similarity is decided.

B. System of Melody Retrieval

Yazawa *et al.*[9] made three consecutive notes one symbol in a melody, and gave the symbol the name. The symbol names are ordered in the melody order. The similarity between the melodies can be defined as the similarity between the character strings of the names. In this way, the similarity between music using the N-gram method. It may be said that it is effective to become the similarity based on the consecutive ups and downs of 3 sounds when we judge whether the moods of music pieces resemble. However, they pay their attention to pitch in their study and did not consider about the length. Therefore, music pieces are judged to be similar if their pitches are similar even if lengths are different. On the other hand, in this study, lengths and pitches are handle as separate factors and calculate the similarity. This enables a similar melody search in consideration of a length.

In addition, Hino *et al.*[10] suggested the similarity calculation using the piano roll image as a new melody technique. This defines the similarity between melodies as the similarity between images where pitch and length information were written down in chronological order from SMFs, which is called a piano roll. It is shown that music pieces having high similarity have highly similar moods. However, as for this technique, it is difficult to use for music pieces whose numbers of notes are very different, because it is greatly affected by the number of notes. The similarity using DTW makes the influence of the number of notes lower when the numbers of notes are greatly different.

III. Methods of Calculating Similarity Between Music Pieces

The aim of this study is to calculate the similarity between an original melody and arrangement, and also calculate it between an arrangement and arrangement with discarding the differences owing to the arrangement. As parts of melodies may resemble while the whole melodies do not resemble, similarity calculation for the whole melodies as well as for the parts of the melodies is proposed.

A. Representation

An SMF has the information of the notes including pitches and length values as codes. These information are expressed as a sequence of vectors $[m(i)] = [(h_i, l_i)]$ in this paper. An element h_i means a pitch value, and l_i means a length value. Sequences of each kind of elements are defined as $[h(i)]$ and $[l(i)]$ because they are separately treated in this paper.

B. Whole Match Distance

The DTW is used in this study. This technique is popular for the similarity calculation technique of time-series data. The DTW is applied to pitches and lengths as time-series data.

There is a problem in applying the DTW to melody information. The DTW is used to measure the distance between melodies. For example, the distance between the values of an original music piece and the music piece created by shifting the original one by one tone becomes large. When this is the one octave shift, the distance calculated becomes remarkably large even if the tones were the same. This is also true for the length. For example, if the note lengths are extended to 1.15 times as compared to the original song, rhythm patterns are exactly the same, while the distance becomes very large. It is considered that the melodies are not similar.

To solve this problem, pre-processing on the melody information is conducted. First, the average pitch value of a time series of pitch values $h = [h(i)]$ is obtained.

$$average(h) = \frac{1}{n}\sum_{i=1}^{n} h(i) \tag{1}$$

Here, n is the number of the notes of the entire melody.

The degree is the difference between the individual $h(i)$ and the $average(h)$.

$$degree(h(i)) = h(i) - average(h) \tag{2}$$

Next, the maximum value of the entire melody $max(h)$ is obtained. The difference $degree(h(i))$ is divided by $max(h)$ to obtain $H(i)$ as shown in Equation (3).

$$H(i) = \frac{degree(h(i))}{max(h)} \times K \tag{3}$$

Here, K(≥ 1) is a constant value to multiply the value since an original value is extremely small. This calculation is applied to the pitch and lengths. It becomes possible to calculate the similarity in consideration with the expansion and contraction of the pitch lag and length.

In addition, the logarithm of this number is also calculated.

$$H^{log}(i) = \frac{\log\{h(i) - average(h)\}}{\log max(h)} \times K \tag{4}$$

When considering the partial match of similarity, this makes the similarity robust against the existence of noise. The result obtained by applying the DTW to the values obtained by Equations (3) and (4) is called $DistanceAll$.

Until here, only a time series of pitch value $h = [h(i)]$ is treated. The same equations are applied to a time series of length values $l = [l(i)]$. In this case, we use $L(i)(L^{log}(i)$, respectively) instead of $H(i)(H^{log}(i))$ in Equation(3)((4)). This makes it possible to normalize the melody information in the time axis.

C. Partial Match Distance

To determine the degree of partial match of two melodies, the sliding window method is used. The slide width is one. Equations (3) and (4) are applied to melody information in the window. The DTW is applied to the values obtained. The expression

$$Slide(Music\,1, WindowSize, SlideWidth)$$

represents the application of the sliding window method to Music 1 with the window width $WindowSize$, and the slide width $SlideWidth$.

The smallest value of the values obtained is the partial match distance: $DistancePart$. The $DistancePart$ between Music 1 and Music 2 is defined as follows:

$$DistancePart = Min(DTW(Music\,1, Slide(Music\,2, WindowSize, 1))) \quad (5)$$

IV. Evaluation

Two types of evaluation experiments are conducted. Validity of Equations (3) and (4) described in Section 3 is examined. Considering partial match distance $DistancePart$ is also examined. The value of K is 100 in this Evaluation.

A. Experimental method

1) Whole Match: In order to examine the validity of Equations (3) and (4), and which is better, 30 MIDI sound sources (No. 1 to No. 30) are created by a professional composer. The average number of notes is 46. They consist of single tones of piano. These have various tempos. In addition, 30 arrangements are made (No. 31 to No. 60) from the original sound sources. Ten arrangements are created by shifting the pitch from -8 to + 8 of No. 1 to No. 10 (No. 31 to No. 40). Other ten are created by varying the whole sound length from 0.75 to 1.09 times from No. 11 to No. 20 (No. 41 to No. 50). Finally, ten arrangements are created by shifting the pitch from -8 to + 8, and lengths 0.75 to 1.09 times of No. 21 to No. 30 (No. 51 to No. 60). After applying Equations (3) and (4) to the total of 60 sound sources, the DTW is applied to the values obtained. Finally, the values are tested through cross-validation.

The computer used in this evaluation is MacBook Pro early 2011 with 2.7 GHz Intel Core i7 (processor) and 4GB RAM. Evaluating program are written in JAVA under OS X Mavericks.

The proposed method is compared with the edit distance (Levenshtein distance). As it is not possible to apply the edit distance to lengths, we carried out distance calculation using a numerical value taking the pitch difference between two adjacent notes. Since this method can not treat the changes in note lengths, only pitches are evaluated.

2) Partial Match: Fifteen materials (No. 61 to No. 75) are created by combining 2 to 4 melodies of sixty original ones (No. 1 to No. 60). Melody combinations used in this experiment are shown in Table 1. For example, the material No. 61 consists of the materials No. 1 and No. 32.

Distances are calculated by using the sliding window methods and Equations (3) and (4) between one of the materials No. 61 to No. 75 and one of the materials No. 1 to No. 60. Here, four window sizes are used. These are 3/4, 1/2, 1/4, and 1/8 of the average number of notes of the material No. 61 to No. 75. The value of K in this experiment is also 100 as used in Experiment 1.

When $DistancePart$ of the corresponding music is the lowest, it is considered that the original song is found. The percentage of the number of the materials found to that of all materials, forty-five is used as the measure of the accuracy of the retrieval. Since 45 melodies are used in this experiment, when all 45 corresponding songs' $DistancePart$ are the lowest, the accuracy becomes 100 %.

TABLE I: Combinations

melody number	1	2	3	4	notes
No. 61	No. 1	No. 32	—	—	99
No. 62	No. 2	No. 33	—	—	101
No. 63	No. 3	No. 34	—	—	105
No. 64	No. 4	No. 35	—	—	113
No. 65	No. 5	No. 36	—	—	108
No. 66	No. 6	No. 37	No. 7	—	173
No. 67	No. 8	No. 38	No. 9	—	220
No. 68	No. 10	No. 39	No. 11	—	190
No. 69	No. 12	No. 40	No. 13	—	177
No. 70	No. 14	No. 41	No. 15	—	172
No. 71	No. 16	No. 42	No. 17	No. 43	170
No. 72	No. 18	No. 44	No. 19	No. 45	169
No. 73	No. 20	No. 46	No. 21	No. 47	144
No. 74	No. 22	No. 48	No. 23	No. 49	166
No. 75	No. 24	No. 50	No. 25	No. 51	136

B. Results

1) Whole Match: Table 2, (Table 3, respectively) shows an example of a result of the distance calculated by using Equation (3) (Equation (4)). In Table 2 (Table 3, respectively), the arrange distance to all of materials (a) ((c)), the distance between the material No. 1 or No. 15 and its arrangement material (B) ((D)), and the ratio of (A) and (B) ((C) and (D)) are shown. Table 2, (Table 3, respectively) does not show the distance of lengths of No. 1 and that of pitches of No.15 because the lengths of arrangement of the No. 1 and pitches of arrangement of the No.15 are the same as those of the original one.

Table 4 and 5 show average distances at each length and each pitch in 3481 sets cross-validated, and all 30 sets of pitches and lengths limited to those between the arrangement and the original one. Then, the results of the methods compared are shown in Table 6.

Figure 1 (2, 3, and 4 respectively) shows the value of the average pitch distance of the entire music was divided by the difference between the original music and arrangement

TABLE II: Results with Equation (3) ($K = 100$)

Target	Average Distance(A)	Distance to the arranged music piece(B)	(A) / (B)
No. 1(pitch)	11.5	0.0297	387.2
No. 15(length)	216.2	0.000119	1810000

TABLE III: Results with Equation (4) ($K = 100$)

Target	Average Distance(C)	Distance to the arranged music piece(D)	(C) / (D)
No. 1(pitch)	0.226	0.0000595	3800
No. 15(length)	1010	7.96	127

TABLE IV: Cross-validation results with the Equation (3) ($K = 100$)

Target	Average Distance(E)	Average between original and arrangement(F)	(E) / (F)
Pitch	8.09	0.0795	101.7
Length	168.5	0.000171	985000

TABLE V: Cross-validation results with the Equation (4) ($K = 100$)

Target	Average Distance(G)	Average between original and arrangement(H)	(G) / (H)
Pitch	0.307	0.000156	1970
Length	1150	28.3	40.6

TABLE VI: Method comparison result for the Music No. 1 ($K = 100$)

compared target	Average Distance(I)	Average between original and arrangement (J)	(I) / (J)
Equation (3)	11.5	0.0297	387.2
Equation (4)	0.226	0.0000595	3800
Edit Distance	49.9	52.0	0.960

distances. This graph shows that the larger the numerical value is, the smaller difference between the original music and arrangement distances are. Because the range of logarithm values is very wide, these are shown in a logarithmic scale in Figure 4.

Figure 5 shows the distance between the original song and the arrangement, the minimum distance of the combinations, the maximum distance, and the average distance.

2) Partial Match: The results are shown in Table 7 (unit: %) . The average processing times per combination are shown in Table 8 in milli seconds.

C. Consideration

1) Whole Match: In the results of Equations (3) and (4), the distance between an original music piece and an arrangement one is significantly lower than the other combinations. The followings have been found from the experiment.

TABLE VII: Accuracy of retrieval [%]

Window Size	Distance of pitch	Logarithm Distance of pitch	Distance of length	Logarithm Distance of length
3/4	11.1	11.1	11.1	13.3
1/2	22.2	16.0	24.4	33.3
1/4	46.4	44.0	40.0	37.7
1/8	26.6	22.2	8.8	28.8

TABLE VIII: Average processing time [msec]

Window Size	No. 61〜 No. 65	No. 66〜 No. 70	No. 71〜 No. 75
3/4	369	996	746
1/2	480	1368	997
1/4	456	1169	863
1/8	343	819	635

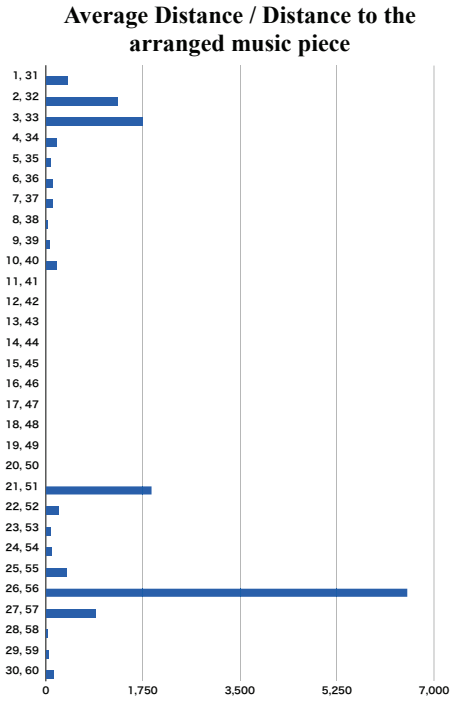

Fig. 1: Pitch Rate
(vertical axis : melody number, horizontal axis : distance)

1) In the method comparison shown in Table 6, in the edit distance, the difference of the average distance between original and arrangement and overall is very close. Then, we can hardly judge whether it is similar or not. However, in each case of Equations (3) and (4), the difference between the total music average distance is more than 100 times of the distance between original and arrangement values. From this fact, in the proposed method, it is easier to find a similar melody.
2) There is no big difference for the pitch in Figures 1 and 2, while the difference can be seen for the note length in Figures 3 and 4. This is because of whether a multiple

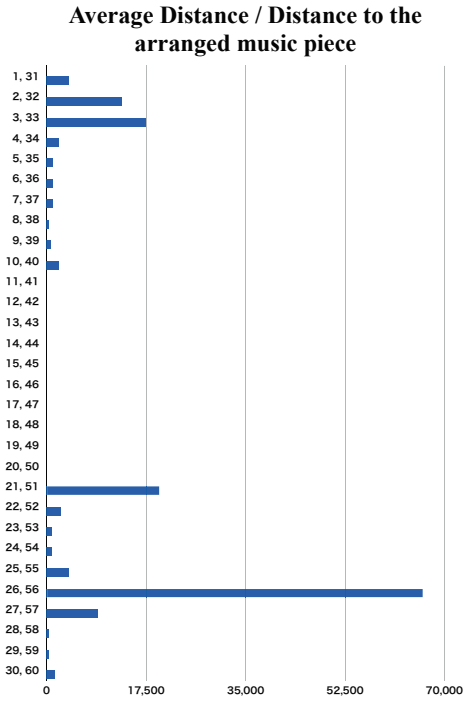

Fig. 2: Logarithm Pitch Rate
(vertical axis : melody number, horizontal axis : distance)

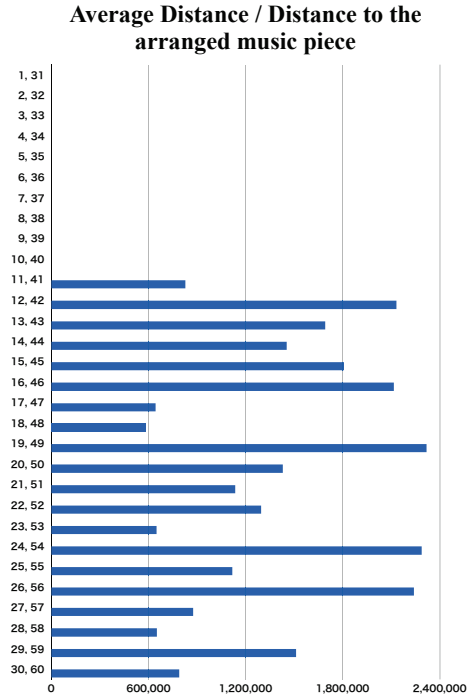

Fig. 3: Length Rate
(vertical axis : melody number, horizontal axis : distance)

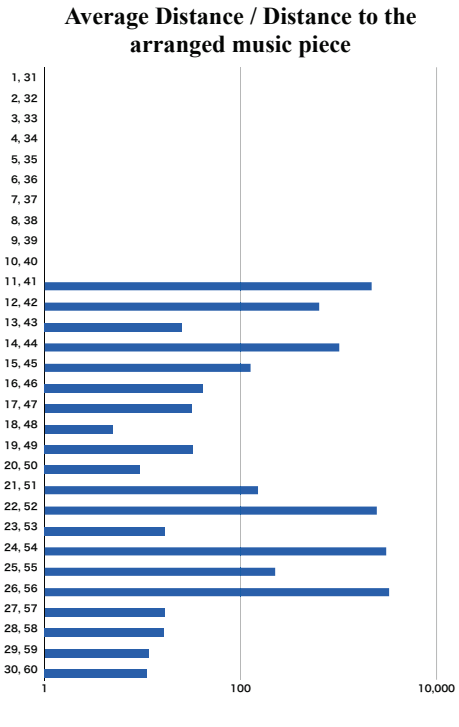

Fig. 4: Logarithm Length Rate
(vertical axis : melody number, horizontal axis : distance)

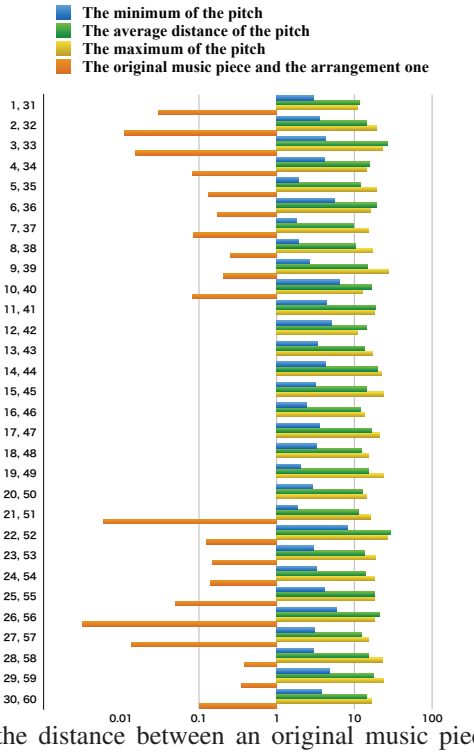

Fig. 5: the distance between an original music piece and an arrangement one, and others
(vertical axis : melody number , horizontal axis : distance)

changes or not. In Figure 4, various note lengths can be seen for the original music pieces No. 11 to No. 30. Only No. 24 has a very high value compared with the others. Conversely, for the other melodies, values are extremely small. The difference of the distances is very small between overall average distance and the original, and the distance between the arrangement. From this result, with respect to the pitch distances, both of Equations (3) and (4) can also be applied to them. For the note length distance, the application of Equation (3) is considered to be good. Therefore, it can be said that it is good to apply Equation (3) to both of pitches and length.

3) From Figure 5, it can be seen that the minimum distances are far apart among the distance between the original and arrangement, and the difference between other combinations. From this result, in the case of introducing a threshold for determining to be similar as a system, the threshold can be set at a power of 10, and only the songs that are actually similar can be obtained, if it is appropriate digits.

2) Partial Match: From Table 7, a significant difference is observed in the accuracy of the window size. The most accurate is 1/4 of the average numbers of the notes of No. 61 to No. 75. The average number of notes of No. 1 to No. 60 is 46. Since it is close to the average number of the notes if the number of notes of No. 66 to No. 75 were divided by four. It is considered that the best accuracy is increased.

Furthermore, from Table 8, as the window size decreases, the processing time tends to decrease. The processing time in 1/2 is longer than 3/4. Calculation time of DTW is proportional to the number of the notes, and the window size. Both are large when the window size is 1/2.

V. Conclusion

We proposed a similarity calculation method considering the similarity between the melodies arranged from the original melody. This is effective in searching melodies arranged because it is possible to significantly reduce the distance between the different melodies whose melody flow are similar.

Extending the proposed method to some useful music formats including MP3 and WAVE is in future work. The distances of the pitch and lengths are independently calculated in this study. It may be required to integrated these distances into one distance. This integration is also in future work. Several different instruments are often used simultaneously. The distance calculation considering the characteristics of instruments is in future work. In this paper, the slide width is set to one. This increases the processing time. Clarifying the appropriate slide width is also in future work. It is also necessary to develop an algorithm to set the appropriate window size.

References

[1] ONKYO ENTERTAINMENT TECHNOLOGY CORPORATION : e-onkyo music, http://www.e-onkyo.com/music/ ,(2015/01/04)

[2] Recochoku, http://recochoku.jp/search_50abc/, (2015/01/04)

[3] J-Lyric.net, http://j-lyric.net/, (2015/01/04)

[4] Utamap, http://www.utamap.com/, (2015/01/04)

[5] Dylan Freedman, Eddie Kohler, Hans Tutschku, " Correlating Extracted and Ground-Truth Harmonic Data in Music Retrieval Tasks", Proceedings of International Conference on Music Information Retrieval (ISMIR) pp. 562-567(2015)

[6] Blair Kaneshiro, Jacek P. Dmochowski, " Neuroimaging Methods for Music Information Retrieval: Current Findings and Future Prospects ", Proceedings of International Conference on Music Information Retrieval (ISMIR) pp. 538-544(2015)

[7] Matev Pesek, Ale Leonardis, Matija Marolt, "A COMPOSITIONAL HIERARCHICAL MODEL FOR MUSIC INFORMATION RETRIEVAL", Proceedings of International Conference on Music Information Retrieval (ISMIR) pp. 131-136(2014)

[8] Peter Grosche, Joan Serrz, Meinard Mller, Josep Lluis Arcos, "STRUCTURE-BASED AUDIO FINGERPRINTING FOR MUSIC RETRIEVAL", Proceedings of 13th International Conference on Music Information Retrieval (ISMIR) pp. 55-60(2012)

[9] Sakurako Yazawa, Masatoshi Hamanaka, "Extension Implication-Realization Model for Subjectivity Melodic Similarity", SIG Report of Information Processing Society of Japan, Vol. 2014-MUS-102, No. 1, pp. 1-5 (2014)

[10] Tatsuya Hino, Taizan Suzuki, Kenzi Noike, Yukio Tokunaga, Kiyoshi Sugiyama, "Similarity Evaluation of Performance Expression by Listening", SIG Report of Information Processing Society of Japan, Vol. 2011-MUS-89, No. 25, pp. 1-6(2011)