Structural Similarity Based on Time-Span Sub-Trees

Masatoshi Hamanaka[1]([✉]), Keiji Hirata[2], and Satoshi Tojo[3]

[1] Kyoto University, Kyoto, Japan
masatosh@kuhp.kyoto-u.ac.jp
[2] Future University Hakodate, Hakodate, Japan
hirata@fun.ac.jp
[3] JAIST, Nomi, Japan
tojo@jaist.ac.jp

Abstract. We propose structural similarity of two melodies based on sub-trees from the time-span tree provided by the Generative Theory of Tonal Music. The structural distance of the tree was previously defined and called the "maximal time-span distance," and experimental results showed to some extent a correspondence between the maximal time-span distance and psychological similarity. However, there is a big problem in that almost all pairs of melodies are not similar on the basis of the maximal time-span distance because the definition of the similarity is too strict. Therefore, we attempt to express a melodic structural similarity by using the coincidence rate of time-span sub-trees to weaken the condition for calculating similarity. We have set up three experimental conditions and compared their results.

Keywords: GTTM · Time-span tree · Similarity · Meet · Join

1 Introduction

We have been developing music analyzers based on the Generative Theory of Tonal Music (GTTM) [1,2]. The main advantage of analysis with GTTM is that it can acquire a tree structure called a "time-span tree" from a score, and this tree structure provides a method for manipulating a piece of music [3].

Before using such manipulation, we have to manually select a melody similar to the target melody. For example, the melody morphing method [3] that generates an intermediate melody between a melody and another melody requires similar melodies for input; otherwise, the two melodies cannot be aligned, and then, a melody cannot be made that has the flavor from both input melodies.

If the values of similarity are not distributed appropriately, it is difficult to represent the search or recommendation result. For example, when there are many pieces of music for which similarity is 1.0 between query pieces, as a result of search or recommendation, there are too many top ranking pieces, and we need a lot of time to preview the top ranking pieces. Therefore, it is desirable

that the similarity value is appropriately distributed without duplication. It is desirable that the obtained subjective similarity results of listening experiments become close to similarity by calculation.

The maximal time-span distance is the sum of the lengths of time spans that match perfectly from root to leaf and is good for measuring the similarity between variations because these variations have a common structure [4]. However, the maximal time-span distance cannot measure the similarity between two different melodies because they usually do not have a common structure.

To measure this similarity, we attempted to weaken the condition for matching time-span trees by using the coincidence rate of time-span sub-trees. We compare three kinds of matching conditions: tight, middle, and weak. Experimental results show that the middle condition is better for the use tasks of searching or recommendation. In the maximal time-span, similarity is calculated by using the sum of matched numbers of notes weighted by the length of the maximal time-span. We also discuss whether this weight is proper or not by considering the subjective similarity and similarity by calculation.

2   Maximal Time-Span Similarity

The time-span tree is a binary tree and is a hierarchical structure describing the relative structural importance of notes that differentiate the essential parts of a melody from the ornamentation. In the tree, the essential notes are connected to a branch nearer to the root of the tree. In contrast, the ornamentation notes are connected to the leaves of the tree. In a separation, we hereafter call the branch "primary" and leaf "secondary" (Fig. 1a). The time-span tree can extract an abstracted melody by reducing ornamentation notes.

Distance between melodies before and after reduction can be expressed as the length of the maximal time-span m_i of a reduced note. Maximal time-span m_i is a sum of time spans t_is of the entire notes, which are recursively connected to the note as secondary. In Fig. 1b, note 2 is connected to note 1 as secondary. Therefore, m_2 is the same as t_2, and m_1 is the sum of t_1 and t_2. In the same way, m_3 is the sum of t_3 and m_1 because note 1 is connected to note 3 as secondary.

We equate melody A with its time-span tree A hereafter. The distance between A and B in Fig. 1b is the sum of maximal time-spans of reduced notes, which we express as $|B - A|$. The maximal time-span distance [4] between P
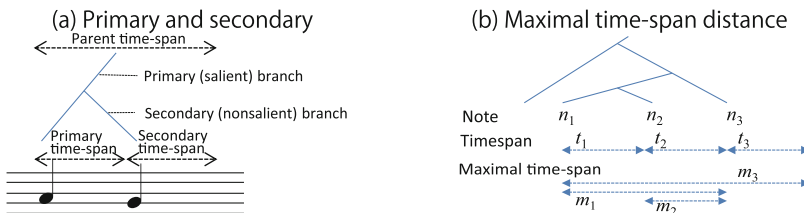

Fig. 1. Time-span trees

and Q can be expressed as the distance via meet $|P - P \sqcap Q| + |Q - P \sqcap Q|$ or distance via join $|P \sqcup Q - P| + |P \sqcup Q - Q|$. The meet operator $P \sqcap Q$ extracts the largest common part or the most common information of the time-span trees in a top-down manner (Fig. 2a). The join operator $P \sqcup Q$ unites two time-span trees in a top-down manner as long as the structures of the two time-span trees are consistent. In fact, the distances via join and meet are the same [4]. The distance between P and Q is maximized when P and Q do not have a common part, which means $P \sqcap Q$ is empty $\perp$ (bottom) as $|P - \perp| + |Q - \perp|$.

The maximal time-span similarity is calculated by normalizing the maximal time-span distance by dividing the maximized maximal time-span distances and subtracting them from 1.[1]

$$1 - \frac{|P - P \sqcap Q|}{2 \cdot |P - \perp|} - \frac{|Q - P \sqcap Q|}{2 \cdot |Q - \perp|} \tag{1}$$

The maximal time-span similarity is higher when the common part of P and Q, which can be expressed as $P \sqcap Q$, is larger. However, the condition for matching is too strict because it compares perfectly from root to leaf through the separation of branches. This strict matching is very good at comparing very similar melodies such as themes and variations [4].

In comparison, the maximal time-span similarity is usually zero between two different melodies.[2] In the experiments, we use time-span trees of 300 8-bar-long monophonic classical music in the GTTM database [5]. In fact, 32 out of 300 pieces in the GTTM database can be interpreted as two kinds of time-span trees. Therefore, we use all the pairs of 332 time-span trees and 32 pairs of the maximal time-span similarity that are bigger than zero.

3   Time-Span Sub-trees Similarity

In the time-span tree, each note is connected to the root through the separations. The matching of maximal time-span similarity is done by comparing all the separations of each note that is a primary or secondary. If the primary/secondary separation was different at some level, all the branches below the separation were discarded as unmatched. In this sub-tree matching, we only pay attention to the final primary/secondary separation, i.e., the separation to leaf note. From now on, we call a pair of a primary branch and a secondary branch with no other separation a *sub-tree*. Figure 2b is an example of matching in the maximal time-span tree, where n_1 is the secondary of the first separation from the root and also the secondary of the second separation from the root, while n_2 is the primary of

---

[1] This definition of similarity is different from general definitions such as of the Jaccard coefficient, the Simpson coefficient, and Dice coefficient in that we independently normalize $P - P \sqcap Q$ and $Q - P \sqcap Q$; otherwise, only one melody is influential when one melody has a lot of notes and the other has little numbers of notes.

[2] The distance by maximal time-span has been improved in [6], however, we here omit its detail by our limited space. We will compare this improved distance by maximal time-span with our sub-tree similarity in future.
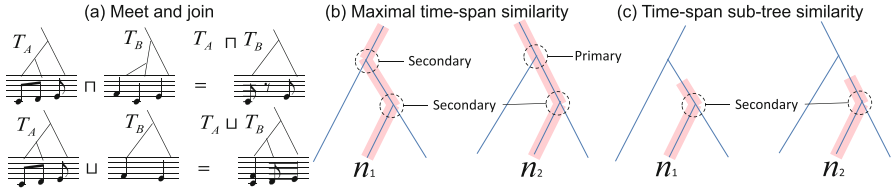
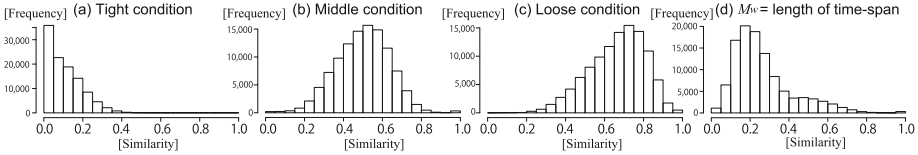**Fig. 2.** Examples of meet ⊓, join ⊔, and matching of nodes

Fig. 3. Histograms of time-span sub-trees similarity

the first separation from the root and secondary of the second separation. Thus, n_1 and n_2 are unmatched. Figure 2c is an example of matching in time-span sub-tree similarity, where n_1 is the secondary of the last separation and also the secondary of last separation. Thus, n_1 and n_2 match.

The matching of maximal time-span similarity is done by comparing the pitches of each note exactly same or not. In our sub-tree matching, we ease the matching condition only focusing on whether two notes of the sub-trees are ascending or descending. Thus, we can regard two sub-trees are matching only if this ascending/descending feature is common between them. Then, we can consider the following three patterns: (i) the pitch of the primary note is higher than that of the secondary, (ii) the pitch of the primary note is lower than that of the secondary, or (iii) the primary and secondary notes have the same pitch. Now we name this classification $M_p = 3$. Furthermore, we can loosen the condition, as (i') the pitch of the primary note is higher than or equal to that of the secondary or (ii') the primary is lower than or equal to the secondary, and we call this classification $M_p = 2$. On the contrary, we can tighten the condition; the pitch of two primary notes must be the same when two sub-trees are matched. We will experiment all these conditions and will compare the results.

The matching of maximal time-spans is done by comparing the onset and length of the maximal time-spans in all separations from the root to leaf that are exactly same. Here we pay attention to the middle point of a maximal time span as there may exist rests at either end of the span. The relative location of the middle points of two time-spans is named *gap*. The matching of time-span sub-tree similarity is done by comparing the gap of the middle of the maximal time-span and the length ratio of the maximal time-span in the sub-trees that are in a certain range. In other words, if the gap of the middle of maximal time-spans is in M_g times of the length of a piece of music and the ratio of the longer maximal time-span and shorter maximal time-span is in M_r, the time-span sub-trees are matching. We will change M_g to 0.1, 0.2, and 0.3 and also change M_r to 1.1, 1.3, and 1.5, and discuss which condition is appropriate.

(a) The nutcracker suite Op.71a
   ``Miniature overture''
   (Pyotr Il'yich Tchaikovsky)

(b) Annen-Polka Op.117
   (Johann Baptist Strauß)

(c) Sonatine Op.20-1 C dur
   (Daniel Friedrich Rudolph Kuhlau)

(d)Sonatine Op.36-5 G dur
   (Muzio Filippo Vincenzo
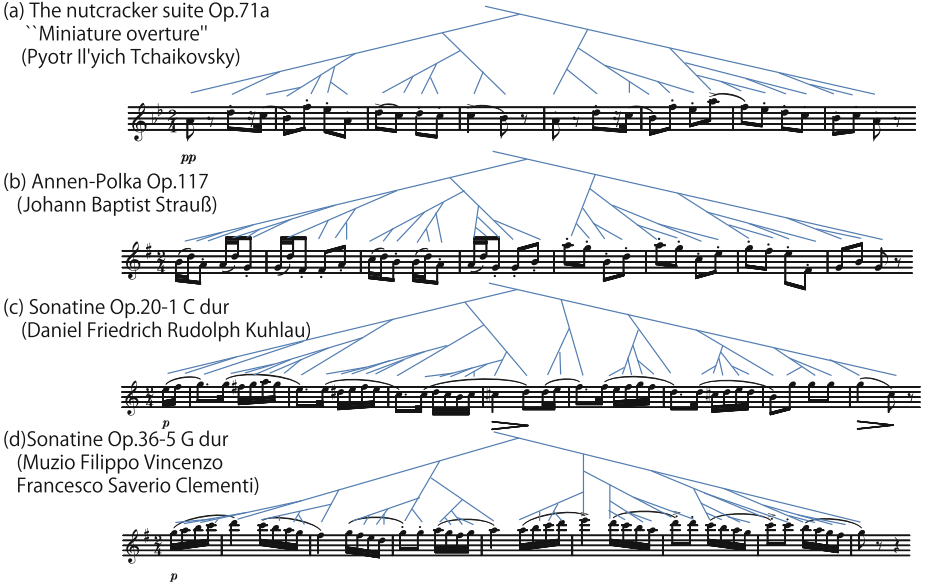   Francesco Saverio Clementi)

Fig. 4. Example of analysis

In maximal time-span similarity, notes close to the root branch are more important than notes far from the root branch because the weights for calculating maximal time-span similarity are the lengths of the maximal time-spans. In the experiment, we compare the calculation of time-span similarity has a weight (M_w = length of the maximal time-span) or has no weight ($M_w = 1$).

4   Experimental Results

The time-span sub-tree similarity proposed in Sect. 3 enables the condition for matching pitch, time, the nodes of the tree, and weight for matiching to be changed. We compare three kinds of matching conditions as follows: (a) Tight condition, with $M_p = 3$, $M_g = 0.1$, $M_r = 1.5$, $M_w = 1$ and exactly match of primary pitch; (b) Middle condition, with $M_p = 3$, $M_g = 0.2$, $M_r = 1.3$, $M_w = 1$; (c) Loose condition, with $M_p = 2$, $M_g = 0.3$, $M_r = 1.5$, $M_w = 1$. Figure 3 is histograms of similarities of each matching condition from the 300 pieces in the database. Compared with the original maximal time-span similarity, Fig. 3a is a decrease in the number of similarities, which is zero; however, most of the similarities are still zero, and all the similarities are under 0.5. In Fig. 3b, the similarities were distributed in a wide area, and the shape of the distribution was similar to the standard distribution. In comparison, the center of the distribution of Fig. 3c leaned to the right, and very few similarities were in the area of zero to 0.3. From the above results, (b) is appropriate in the three conditions.

We compared two kinds of weight for matching: (c) $M_w = 1$ and (d) $M_w =$ the length of maximal time-span. Here, for the values of M_p, M_g, and M_r, we

use the condition of (b). Figure 3d is a histogram of similarity made by using the weight as (d), which has also a long distribution.

As described in Sect. 2, the 32 out of 300 pieces in the GTTM database have two kinds of time-span trees corresponding to interpretation from musicologists. It is preferable that each two kinds of time-span trees are similar because there are interpretations of the same piece. We calculated the average similarity of the 32 pieces by using the weights of (c) and (d). As a result, the average of similarity for (c) was 0.77 and for (d) was 0.6. To compare the values from (c) and (d), we normalized the average to 0 and variance to 1 because the average and distribution of similarities from (c) and (d) were different. As a result of normalization, the average similarity for (c) was 2.16 and for (d) was 2.65. These values depend on the character of the corpus. In current 300 piece the weight for matching is appropriately by using condition of (c) than (d).

We show some examples that show high similarities where the condition is (b). The similarities of Fig. 4a and b was 0.90, Fig. 4c and d was 0.88.

5   Conclusion

Although maximal time-span similarity was previously proposed, the similarity of two different pieces of music is usually zero. We proposed time-span sub-tree similarity, for which similarity is made to be more than zero in many cases by weakening the condition for matching maximal time-span similarity. In the experimental results obtained with three kinds of matching condition, the middle one was better than the other conditions with the GTTM database. We plan to construct applications for similarity with the time-span tree because the appropriate definition of similarity depends on the application.

References

1. Lerdahl, F., Jackendoff, R.: A Generative Theory of Tonal Music. MIT Press, Cambridge (1983)
2. Hamanaka, M., Hirata, K., Tojo, S.: Implementing "A generative theory of tonal music". J. New Music Res. **35**(4), 249–277 (2007)
3. Hamanaka, M., Hirata, K., Tojo, S.: Melody morphing method based on GTTM. In: Proceeding of ICMC 2008, Belfast, pp. 155–158 (2008)
4. Tojo, S., Hirata, K.: Structural similarity based on time-span tree. In: Aramaki, M., Barthet, M., Kronland-Martinet, R., Ystad, S. (eds.) CMMR 2012. LNCS, vol. 7900, pp. 400–421. Springer, Heidelberg (2013)
5. Hamanaka, M., Hirata, K., Tojo, S.: Music structural analysis database based on GTTM. In: Proceeding of ISMIR 2014, Taipei, pp. 325–330 (2014)
6. Hirata, K., Tojo, S., Hamanaka, M.: Cognitive similarity grounded by tree distance from the analysis of K.265/300e. In: Aramaki, M., Derrien, O., Kronland-Martinet, R., Ystad, S. (eds.) CMMR 2013. LNCS, vol. 8905, pp. 589–605. Springer, Heidelberg (2014)