# Cognitive Similarity Grounded by Tree Distance from the Analysis of K.265/300e

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

[1] Future University Hakodate, 116-2 Kamedanakano-cho, Hakodate, Hokkaido 041-8655, Japan
hirata@fun.ac.jp
[2] Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
tojo@jaist.ac.jp
[3] Kyoto University, 46-29 Yoshida-Shimoadachicho, Sakyo, Kyoto 606-8501, Japan
masatosh@kuhp.kyoto-u.ac.jp

**Abstract.** Lerdahl and Jackendoff's theory employed a tree in a representation of internal structure of music. In order for us to claim that such a tree is a consistent and stable representation, we argue that the difference of trees should correctly reflect our cognitive distance of music. We report our experimental result concerning the comparison of similarity among variations on *Ah vous dirai-je, maman*, K. 265/300e by Mozart. First we measure the distance in trees between two variations by the sum of the lengths of time-spans, and then we compare the result with the human psychological similarity. We show a statistical analysis of the distance and discuss the adequacy of it as a criteria of similarity.

**Keywords:** Time-span tree · Generative theory of tonal music · Join/meet operations · Cognitive similarity

## 1 Introduction

Music theory gives us methodology to analyze music written on scores, and clarifies their inherent features in a comprehensive way. There have been many attempts to embody a music theory onto a computer system and to build a music analyzer. In particular, some music theories employ trees to represent the deep structure of a musical piece [1,6,10,11,15], and such a tree representation seems a promising way to automatize the analyzing process. It is, however, widely recognized that there are intrinsic difficulties in this; (i) how we can formalize ambiguous or missing concepts and (ii) how we can assess the consistency and stability of a fomalized music theory.

For (i), the approaches include the externalization of those hidden features of music. For example, Lerdahl and Jackendoff [10] (the generative theory of tonal music; GTTM hereafter) specified many rules to retrieve such information in music to obtain a time-span tree, though they proposed only heuristics


M. Aramaki et al. (Eds.): CMMR 2013, LNCS 8905, pp. 589–605, 2014.
DOI: 10.1007/978-3-319-12976-1_36

and missed fully explicit algorithms. Thus, to formalize this theory, we have complemented necessary parameters to clarify the process of time-span reduction [8]. Pearce and Wiggins [14] have built a model to derive as many features as possible from the scores; these features contain the properties of potentially non-contiguous events.

For (ii), we do not think there is an agreeable solution yet, but we propose to assess the consistency and the stability of a formalized music theory based on our cognitive reality. If the tree representation derived by a formalized music theory is sufficiently stable and consistent, the distance between those representations must reflect our human intuition on difference in music. Hence, if we are to compare the tree distance and our psychological difference and/or similarity, we can evaluate the consistency and the stability of a formalized music theory.

Here, we look back at the studies on similarity in music. In music information research, the similarity has been drawing attention of many researchers [9,19]. Some of the researchers are motivated by engineering demands such as music retrieval, classification, and recommendation, [7,13,16] and others by modeling the cognitive processes of musical similarity [4,5]. Several types of similarity have been proposed, including melodic similarity, e.g., van Kranenburg (2010) [18] and harmonic similarity, e.g., de Haas (2012) [3]. The song similarity in MIREX is widely recognized as an important category in the contest [12]. All these viewpoints suggest the importance of quantitative comparison, and thus we employ a numeric distance in measuring the cognitive similarity.

We have proposed a notion of distance among the time-span trees [17], however, in that research there lacked the discussion on the perception of similarity. As a result, it was difficult for us to explain that the distance could be a metric of similarity [3,9,13,18]. The contribution of this paper is that we have actually conducted a psychological experiment on the similarity among 12 variations on *Ah vous dirai-je, maman*, K. 265/300e by Wolfgang Amadeus Mozart, in comparison with the corresponding time-span distance.

This paper is organized as follows: in Sect. 2 we briefly summarize the notion of time-span tree and reduction. In Sect. 3, we introduce our notion of distance in time-span trees. Then, to apply the notion to arbitrary two music pieces, we generalize the distance in Sect. 4. In Sect. 5, we report our experimental result. First we measure the distance between two variations by the tree distance, and then we compare the result with the human psychological resemblance. We show the statistical analysis, and discuss the adequacy of our measure as a metric of similarity in Sect. 6. Finally we conclude in Sect. 7.

## 2 Time-Span Tree and Reduction

Time-span reduction in Lerdahl and Jackendoff's Generative Theory of Tonal Music (GTTM; hereafter) [10] assigns structural importance to each pitch events in the hierarchical way. The structural importance is derived from the *grouping analysis*, in which multiple notes compose a short phrase called a group, and from the *metrical analysis*, where strong and weak beats are properly assigned on each

pitch event. As a result, a time-span tree becomes a binary tree constructed in bottom-up and top-down manners by comparison between the structural importance of adjacent pitch events at each hierarchical level. Although a pitch event means a single note or a chord, we restrict our interest to monophonic analysis in this paper, as the method of chord recognition is not included in the original theory. Figure 1 shows an excerpt from [10] demonstrating the concept of reduction. In the sequence of reductions, each reduction should sound like a simplification of the previous one. In other words, the more reductions proceed, each sounds dissimilar to the original. Reduction can be regarded as abstraction, but if we could find a proper way of reduction, we can retrieve a basic melody line of the original music piece. The key idea of our framework is that reduction is

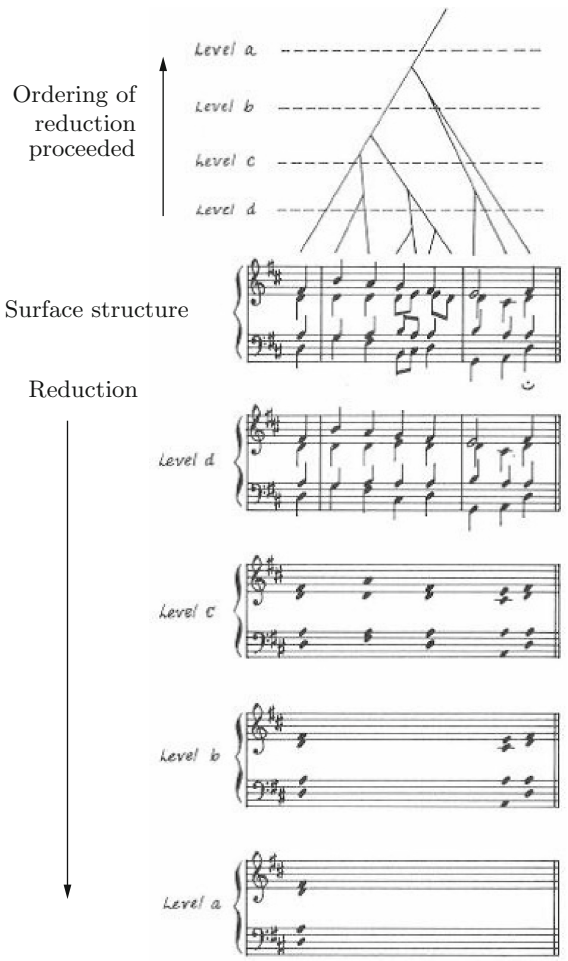



**Fig. 1.** Reduction hierarchy of chorale 'O Haupt voll Blut und Wunden' in St. Matthew's Passion by J.S. Bach [10, p. 115]

identified with the subsumption relation, which is the most fundamental relation in knowledge representation.

## 3 Strict Distance in Time-Span Reduction

In the section, the basic formalization of time-span trees is given as a prerequisite for extending our framework to be described later. Since the contents presented in the section overlap some of those in [17] and contain rather mathematical stuff, the readers who first want to comprehend an outline of the contributions could move onto the experimental section. Then, the readers would come back here afterward.

### 3.1 Subsumption, *Join*, and *Meet*

First we define the notion of *subsumption*. Let $\sigma_1$ and $\sigma_2$ be tree structures. $\sigma_2$ subsumes $\sigma_1$, that is, $\sigma_1 \sqsubseteq \sigma_2$ if and only if for any branch in $\sigma_1$ there is a corresponding branch in $\sigma_2$.

**Definition 1 (*Join* and *Meet*).** *Let $\sigma_A$ and $\sigma_B$ be tree structures for music $A$ and $B$, respectively. If we can fix the least upper bound of $\sigma_A$ and $\sigma_B$, that is, the least $y$ such that $\sigma_A \sqsubseteq y$ and $\sigma_B \sqsubseteq y$ is unique, we call such $y$ the join of $\sigma_A$ and $\sigma_B$, denoted as $\sigma_A \sqcup \sigma_B$. If we can fix the greatest lower bound of $\sigma_A$ and $\sigma_B$, that is, the greatest $x$ such that $x \sqsubseteq \sigma_A$ and $x \sqsubseteq \sigma_B$ is unique, we call such $x$ the meet of $\sigma_A$ and $\sigma_B$, denoted as $\sigma_A \sqcap \sigma_B$.*

We illustrate *join* and *meet* in a simple example in Fig. 2. The '$\sqcup$' (*join*) operation takes quavers in the scores, so that missing note in one side is complemented. On the other hand, the '$\sqcap$' (*meet*) operation takes $\bot$ for mismatching features [2], and thus only the common notes appear as a result.

Obviously from Definition 1, we obtain the absorption laws: $\sigma_A \sqcup x = \sigma_A$ and $\sigma_A \sqcap x = x$ if $x \sqsubseteq \sigma_A$. Moreover, if $\sigma_A \sqsubseteq \sigma_B$, $x \sqcup \sigma_A \sqsubseteq x \sqcup \sigma_B$ and $x \sqcap \sigma_A \sqsubseteq x \sqcap \sigma_B$ for any $x$.

We can define $\sigma_A \sqcup \sigma_B$ and $\sigma_A \sqcap \sigma_B$ in recursive functions. In the process of unification between $\sigma_A$ and $\sigma_B$, when a single branch is unifiable with a tree,

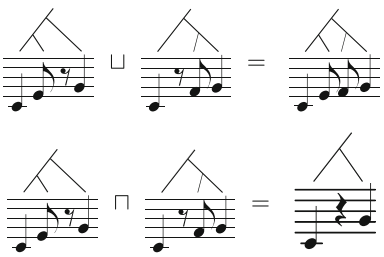


**Fig. 2.** *Join* and *meet*

$\sigma_A \sqcup \sigma_B$ chooses the tree while $\sigma_A \sqcap \sigma_B$ chooses the branch, in a recursive way. Because there is no alternative action in these procedures, $\sigma_A \sqcup \sigma_B$ and $\sigma_A \sqcap \sigma_B$ exist uniquely. Thus, the partially ordered set of time-span trees becomes a *lattice*.

### 3.2 Maximal Time-Span and Reduction Distance

In GTTM, a listener is supposed to construct mentally pitch hierarchies (reductions) that express maximal importance among pitch relations [10, p. 118]. We here observe a time-span becomes longer as the level of time-span hierarchy goes higher. Then, we can suppose that a longer time-span contains more information, and it is therefore regarded more important.

Based on the above consideration, we hypothesize:

> *If a branch with a single pitch event is reduced, the amount of information corresponding to the length of its time-span is lost.*

We call a sequence of reductions of a music piece *reduction path*. We regard the sum of the length of such lost time-spans as the distance of two trees, in the reduction path. Thereafter, we generalize the notion to be feasible, not only in a reduction path but in any direction in the lattice.

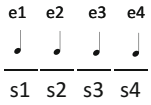
We presuppose that branches are reduced only one by one, for the convenience to sum up distances. A branch is *reducible* only in the bottom-up way, i.e., a reducible branch possesses no other sub-branches except a single pitch event at its leaf. In the similar way, we call the reverse operation *elaboration*; we can attach a new sub-branch when the original branch consists only of a single event.

The *head* pitch event of a tree structure is the most salient event of the whole tree. Though the event itself retains its original duration, we may regard its saliency is extended to the whole tree. The situation is the same as each subtree. Thus, we consider that each pitch event has the maximal length of saliency.

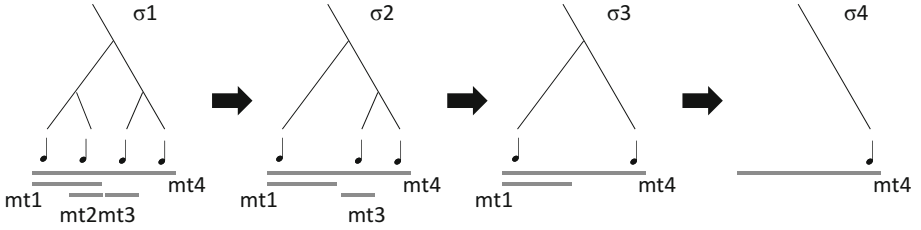
**Definition 2 (Maximal Time-Span).** *Each pitch event has the maximal time-span within which the event becomes most salient, and outside the time-span the salience is lost.*

In Fig. 3(a), there are four contiguous pitch events, e1, e2, e3, and e4; each has its own temporal span (duration on surface), s1, s2, s3, and s4, denoted thin lines. Figure 3(b) depicts time-span trees and corresponding maximal time-span hierarchies, denoted gray thick lines. The relationships between spans in (a) and maximal time-spans in (b) as follows. At the lowest level in the hierarchy, the length of a span is equal to that of a maximal time-span; $\mathrm{mt2} = \mathrm{s2}$, $\mathrm{mt3} = \mathrm{s3}$. At the higher levels, $\mathrm{mt1} = \mathrm{s1} + \mathrm{mt2}$, and $\mathrm{mt4} = \mathrm{mt1} + \mathrm{mt3} + \mathrm{s4} = \mathrm{s1} + \mathrm{s2} + \mathrm{s3} + \mathrm{s4}$. That is, every span extends itself by concatenating the span at a lower level along the configuration of a time-span tree. When all subordinate spans are concatenated up into a span, the span reaches the maximal time-span.

Only the events at the lowest level in the hierarchy are reducible; the other events cannot be reduced. In Fig. 3(b), for the leftmost time-span tree $\sigma 1$, either




(a) Sequence of pitch events and their spans




(b) Reduction proceeds by removing a reducible maximal time-span

**Fig. 3.** Reduction of time-span tree and maximal time-span hierarchy; gray thick lines denote maximal time-spans while thin ones pitch durations

e2 or e3 is reducible; e2 is first reduced then e3. For $\sigma 2$, e3 is only reducible, not e1 because e1 is not at the lowest level in the maximal time-span hierarchy.

Let $\varsigma(\sigma)$ be a set of pitch events in $\sigma$, $\sharp\varsigma(\sigma)$ be its cardinality, and $s_e$ be the maximal time-span of event $e$. Since reduction is made by one reducible branch at a time, a reduction path $\sigma^n, \sigma^{n-1}, \ldots, \sigma^2, \sigma^1, \sigma^0$, such that $\sigma^n \sqsupseteq \sigma^{n-1} \sqsupseteq \cdots \sqsupseteq \sigma^2 \sqsupseteq \sigma^1 \sqsupseteq \sigma^0$, suffices $\sharp\varsigma(\sigma^{i+1}) = \sharp\varsigma(\sigma^i) + 1$. If we put $\sigma_A = \sigma^0$ and $\sigma_B = \sigma^n$, $\sigma_A \sqsubseteq \sigma_B$ holds by transitivity. For each reduction step, when a reducible branch on event $e$ disappears, its maximal time-span $s_e$ is accumulated as distance.

**Definition 3 (Reduction Distance).** *The distance $d_{\sqsubseteq}$ of two time-span trees such that $\sigma_A \sqsubseteq \sigma_B$ in a reduction path is defined by*

$$d_{\sqsubseteq}(\sigma_A, \sigma_B) = \sum_{e \in \varsigma(\sigma_B) \setminus \varsigma(\sigma_A)} s_e.$$

For example in Fig. 3, the distance between $\sigma 1$ and $\sigma 4$ becomes $\mathrm{mt1} + \mathrm{mt2} + \mathrm{mt3}$. Note that if e3 is first reduced and e2 is subsequently reduced, the distance is the same. Although the distance is a simple summation of maximal time-spans at a glance, there is a latent order in the addition, for reducible branches are different in each reduction step. In order to give a constructive procedure on this summation, we introduce the notion of total sum of maximal time-spans.

**Definition 4 (Total Maximal Time-Span).** *Given tree structure $\sigma$,*

$$tmt(\sigma) = \sum_{e \in \varsigma(\sigma)} s_e.$$

When $\sigma_A \sqsubseteq \sigma_B$, from Definitions 3 and 4, $d_{\sqsubseteq}(\sigma_A, \sigma_B) = tmt(\sigma_B) - tmt(\sigma_A)$. As a special case of the above, $d_{\sqsubseteq}(\bot, \sigma) = tmt(\sigma)$.

### 3.3 Properties of Distance

*Uniqueness of Reduction Distance:* First, as there is a reduction path between $\sigma_A \sqcap \sigma_B$ and $\sigma_A \sqcup \sigma_B$, and $\sigma_A \sqcap \sigma_B \sqsubseteq \sigma_A \sqcup \sigma_B$, $d_{\sqsubseteq}(\sigma_A \sqcap \sigma_B, \sigma_A \sqcup \sigma_B)$ is computed by the difference of total maximal time-span. Because the algorithm returns a unique value, for any reduction path from $\sigma_A \sqcup \sigma_B$ to $\sigma_A \sqcap \sigma_B$, $d_{\sqsubseteq}(\sigma_A \sqcap \sigma_B, \sigma_A \sqcup \sigma_B)$ is unique. This implies the uniqueness of reduction distance: if there exist reduction paths from $\sigma_A$ to $\sigma_B$, $d_{\sqsubseteq}(\sigma_A, \sigma_B)$ is unique.

Next, from set-theoretical calculus, $\varsigma(\sigma_A \sqcup \sigma_B) \backslash \varsigma(\sigma_A) = \varsigma(\sigma_B) \backslash \varsigma(\sigma_A \sqcap \sigma_B)$. Then, $d_{\sqsubseteq}(\sigma_A, \sigma_A \sqcup \sigma_B) = \sum_{e \in \varsigma(\sigma_A \sqcup \sigma_B) \backslash \varsigma(\sigma_A)} s_e = \sum_{e \in \varsigma(\sigma_B) \backslash \varsigma(\sigma_A \sqcap \sigma_B)} s_e = d_{\sqsubseteq}(\sigma_A \sqcap \sigma_B, \sigma_B)$. Therefore, $d_{\sqsubseteq}(\sigma_A, \sigma_A \sqcup \sigma_B) = d_{\sqsubseteq}(\sigma_A \sqcap \sigma_B, \sigma_B)$ and $d_{\sqsubseteq}(\sigma_B, \sigma_A \sqcup \sigma_B) = d_{\sqsubseteq}(\sigma_A \sqcap \sigma_B, \sigma_A)$.

Here let us define two ways of distances.

$$d_{\sqcap}(\sigma_A, \sigma_B) = d_{\sqsubseteq}(\sigma_A \sqcap \sigma_B, \sigma_A) + d_{\sqsubseteq}(\sigma_A \sqcap \sigma_B, \sigma_B)$$
$$d_{\sqcup}(\sigma_A, \sigma_B) = d_{\sqsubseteq}(\sigma_A, \sigma_A \sqcup \sigma_B) + d_{\sqsubseteq}(\sigma_B, \sigma_A \sqcup \sigma_B)$$

Then, we immediately obtain $d_{\sqcup}(\sigma_A, \sigma_B) = d_{\sqcap}(\sigma_A, \sigma_B)$ by the uniqueness of reduction distance.

For any $\sigma', \sigma''$ such that $\sigma_A \sqsubseteq \sigma' \sqsubseteq \sigma_A \sqcup \sigma_B$, $\sigma_B \sqsubseteq \sigma'' \sqsubseteq \sigma_A \sqcup \sigma_B$, $d_{\sqcup}(\sigma_A, \sigma') + d_{\sqcap}(\sigma', \sigma'') + d_{\sqcup}(\sigma'', \sigma_B) = d_{\sqcup}(\sigma_A, \sigma_B)$. Ditto for the meet distance. Now the notion of distance, which was initially defined in the reduction path as $d_{\sqsubseteq}$ is now generalized to $d_{\{\sqcap,\sqcup\}}$, and in addition we have shown they have the same values. From now on, we omit $\{\sqcap, \sqcup\}$ from $d_{\{\sqcap,\sqcup\}}$, simply denoting '$d$'. Here, $d(\sigma_A, \sigma_B)$ is unique among shortest paths between $\sigma_A$ and $\sigma_B$. Note that shortest paths can be found in ordinary graph-search methods, such as *branch and bound*, Dijkstra's algorithm, best-first search, and so on. As a corollary, we also obtain $d(\sigma_A, \sigma_B) = d(\sigma_A \sqcup \sigma_B, \sigma_A \sqcap \sigma_B)$.

*Triangle Inequality:* Finally, as $d(\sigma_A, \sigma_B) + d(\sigma_B, \sigma_C)$ becomes the sum of maximal time-spans in $\varsigma(\sigma_A \sqcup \sigma_B) \backslash \varsigma(\sigma_A \sqcap \sigma_B)$ plus those in $\varsigma(\sigma_B \sqcup \sigma_C) \backslash \varsigma(\sigma_B \sqcap \sigma_C)$ while $d(\sigma_A, \sigma_C)$ becomes $\varsigma(\sigma_A \sqcup \sigma_C) \backslash \varsigma(\sigma_A \sqcap \sigma_C)$, we obtain $d(\sigma_A, \sigma_B) + d(\sigma_B, \sigma_C) \geq d(\sigma_A, \sigma_C)$: the triangle inequality. For more details on the theoretical stuff, see [17].

In Fig. 4, we have laid out various reductions originated from a piece. As we can find three reducible branches in $A$ there are three different reductions: $B$, $C$, and $D$. In the figure, $C$ (shown diluted) lies behind the lattice where three back-side edges meet. The distances, represented by the length of edges, from $A$ to $B$, $D$ to $F$, $C$ to $E$, and $G$ to $H$ are the same, since the reduced branch is common. Namely, the reduction lattice becomes parallelepiped,[1] and the distances from $A$ to $H$ becomes uniquely $2 + 2 + 2 = 6$. We exemplify the triangle inequality; from $A$ through $B$ to $F$, the distance becomes $2+2 = 4$, and that from $F$ through $D$ to $G$ is $2 + 2 = 4$, thus the total path length becomes $4 + 4 = 8$. But, we can find a shorter path from $A$ to $G$ via either $C$ or $D$, in which case the distance

[1] In the case of Fig. 4, as all the edges have the length of 2, the lattice becomes equilateral.

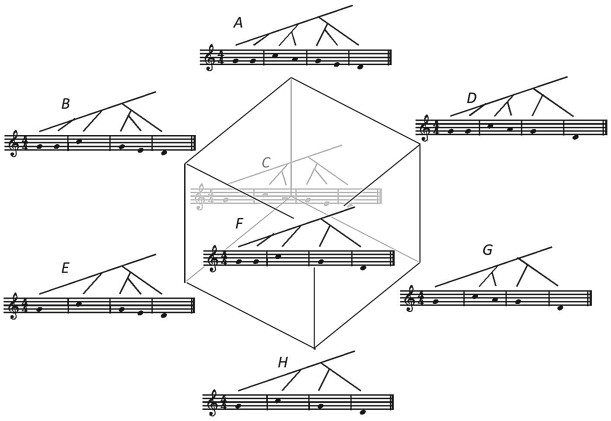



**Fig. 4.** Reduction lattice

becomes $2 + 2 = 4$. Notice that the lattice represents the operations of *join* and *meet*; e.g., $F = B \sqcap D$, $D = F \sqcup G$, $H = E \sqcap F$, and so on. In addition, the lattice is locally Boolean, being $A$ and $H$ regarded to be $\top$ and $\bot$, respectively. That is, there exists a complement,[2] and $E^c = D$, $C^c = F$, $B^c = G$, and so on.

Notice that time-span tree includes more information than a real score, as head and grouping hierarchy is added. On the other hand, a real score contains the information of each note and rest (e.g., onset and duration), which is lost in the time-span tree. In this sense, for a time-span tree, *rendering* procedure[3] needs to be compensated, to be an audible music. For example, let us consider level c in Fig. 1 which contains five pitch events (chords), the second of which is made of notes F♯5 A5 D6 and A6 drawn at the second beat of the first bar as if they would sound at the position. However, the maximal time-span of the chord begins at the first beat of the first bar and sustains for the duration of a half note. Therefore, we need to take care not to confuse the position of time-spans from those of rendered pitch events. Actually, in [10] the authors have avoided to add stems to note heads upon drawing individual notes in the reduction process in Fig. 1.

## 4 Generalized Distance in Trees

In this section, we extend the notion of strict distance, to be applicable to two different music pieces, which may not necessarily share a common-ancestor music piece in terms of reduction. To this purpose, we need to relax the condition of distance calculation.

[2] For any member $X$ of a set, there exists $X^c$ and $X \sqcup X^c = \top$ and $X \sqcap X^c = \bot$.

[3] Rendering was originally introduced in computer graphics, which means the operation of creating images from a model, such as a photorealistic image from a wireframe model.

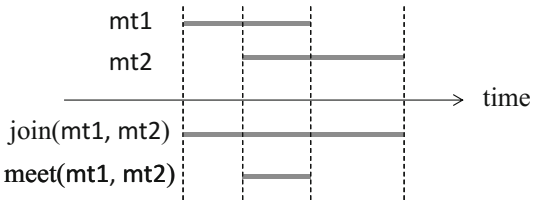



**Fig. 5.** Generalized join and meet operations to maximal time-spans

### 4.1 Interval Semantics on Absolute Time Axis

In order to compare two different melodies, we need to place those at proper places in a common temporal axis. Two arbitrary music pieces are possibly different from each other with a large variety; for example, a music piece beginning with auftakt or syncopation, containing hemiola, and being at a double tempo with the same pitch sequence. To handle such cases, we may need various types of adjustments of two music pieces for comparison; for example, alignment by the endpoints of music pieces and/or bar lines, and by stretching or compressing to make the two of same length.

At present, we take the simplest approach to the adjustment in which two music pieces are aligned only at the beginning bar line. Then, the join/meet operations are applied to maximal time-spans without stretching or compressing them. That is, when two temporal intervals have a common length, the result of a join operation encompasses the temporal union of the two intervals, and that of meet operation is exactly the temporal intersection (Fig. 5), where $\mathrm{mt}\{1, 2\}$ means a maximal time-span, respectively. The decision is underlain by the following assumption: the longer a time-span is, the more informative it is, as the interval semantics of temporal logic [21].

### 4.2 Meet-Oriented Distance

In Sect. 3.3, we have shown that the distance via the meet and that via the join become the same in the lattice of strict descendents of one common music piece. However, when we are to apply two music pieces without a common ancestor, one serious problem is that such equality of join/meet distance may not be promised. First of all, we cannot calculate the join operation for all cases at present; for example, if the supremacy of the heads of two trees do not match, the result of the join operation is not defined (Fig. 6). In contrast, the result of meet operation can be calculated in any case. Therefore, in this paper, we decide to calculate the distance using the path via the meet $d_{\sqcap}$ in Sect. 3.3.

Figure 7 shows the excerpt from the Prolog program implementing the generalized join and meet operations. The join/meet operations are recursively applied in the top-down manner. In the Prolog program, a node in a time-span tree is represented by data structure $(T_p > T_s)$ or $(T_s > T_p)$, where $T_p$ and $T_s$ denote subtrees (Fig. 8). Subscripts 'p' and 's' represent that a branch is primary or secondary, and the temporal order between them is shown by '<' or '>'.

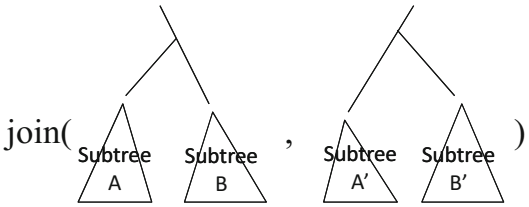



**Fig. 6.** Case of undefined result in join operation

```
join(X,Y,Join) :-
X = (Xp > Xs),
Y = (Yp > Ys),!,
join(Xp,Yp,Mp),
join(Xs,Ys,Ms),
Join = (Mp > Ms).
join(X,Y,Join) :-
X = (Xs < Xp),
Y = (Ys < Yp),!,
join(Xp,Yp,Mp),
join(Xs,Ys,Ms),
Join = (Ms < Mp).
join(X,Y,Join) :-
X = (_ > _),
Y = (_ < _),!,
Join = undefined.
join(X,Y,Join) :-
X = (_ < _),
Y = (_ > _),!,
Join = undefined.
```

```
meet(X,Y,Meet) :-
X = (Xp > Xs),
Y = (Yp > Ys),!,
meet(Xp,Yp,Mp),
meet(Xs,Ys,Ms),
Meet = (Mp > Ms).
meet(X,Y,Meet) :-
X = (Xs < Xp),
Y = (Ys < Yp),!,
meet(Xp,Yp,Mp),
meet(Xs,Ys,Ms),
Meet = (Ms < Mp).
meet(X,Y,Meet) :-
X = (Xp > _),
Y = (_ < Yp),!,
meet(Xp,Yp,Meet).
meet(X,Y,Meet) :-
X = (_ < Xp),
Y = (Yp > _),!,
meet(Xp,Yp,Meet).
```

**Fig. 7.** Prolog implementation of generalized join and meet operations (recursion part)

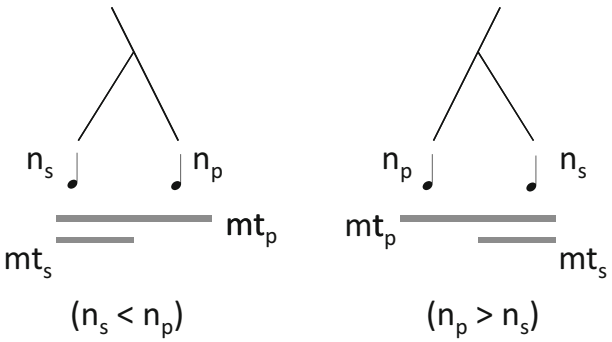



**Fig. 8.** Representation of time-span tree node in Prolog program

For treating the unmatched supremacy of the heads of two trees, the result of join is set to `undefined` in the third and fourth clauses of `join` in Fig. 7, and that of meet is calculated by the further meet of primary branches, `Xp` and `Yp`, ignoring secondary branches, in the third and fourth clauses of `meet`.

## 5 Experiment

We conduct two experiments using the same set of pieces: a similarity assessment by human listeners and the calculation by the proposed framework. Set piece is the Mozart's variations K.265/300e '*Ah, vous dirai-je, maman*', also known as '*Twinkle, twinkle little star*'. The piece consists of the famous theme and twelve variations of it. In our experiment, we excerpt the first eight bars (Fig. 9). Although the original piece includes multiple voices, our framework can only treat monophony; therefore, the original piece is arranged into a monophony. We extract salient pitch events from each one of two voices, choosing a prominent note from a chord, and disregard the difference of octave so that the resultant melody is heard smoothly. In total, we have the theme and twelve variations (eight bars long) and obtain 78 pairs to be compared (${}_{13}C_2 = 78$).

For the similarity assessment by human listeners, eleven university students participate in our study, seven out of whom have experiences in playing music instruments more or less. An examinee listens to all pairs $\langle m_1, m_2 \rangle$ in the random order without duplication, where $m_{\{1,2\}}$ is either themeor variations No. 1–12. Every time he/she listens to it, he/she is asked "how similar is $m_1$ to $m_2$?", and ranks it in one of five grades among *quite similar* $= 2$, *similar* $= 1$, *neutral* $= 0$, *not similar* $= -1$, and *quite different* $= -2$. At the very beginning, for cancelling the cold start bias, every examinee hears the theme and twelve variations (eight bars long) through without ranking them. In addition, when an examinee listens to and rank pair $\langle m_1, m_2 \rangle$, he/she should try the same pair later to avoid the order effect. Finally, the average rakings are calculated within an examinee and then for all the examinees.

For the calculation by the proposed framework, we use the meet-oriented distance introduced in Sect. 4.2. From Definitions 3 and 4, the distance is measured by a note duration, we set the unit of distance to one third of the sixteenth note duration so that a music piece not only in quadruple time but also in triple time can be represented. The correct time-span trees of the theme and twelve variations are first created by the authors and are next cross-checked to each other. Note that the meet operation takes into account only the configuration of a time-span tree, not pitch events; it is obvious from Definitions 3 and 4.

## 6 Results and Analysis

The experimental results are shown in the distance-matrix (Table 1). The theoretical estimation (a) means the results of calculation by the meet operation, and the human listeners (b) means the psychological resemblance by examinees. In (a), since the values of *meet*($m_1$,$m_2$) and *meet*($m_2$,$m_1$) are exactly the same,

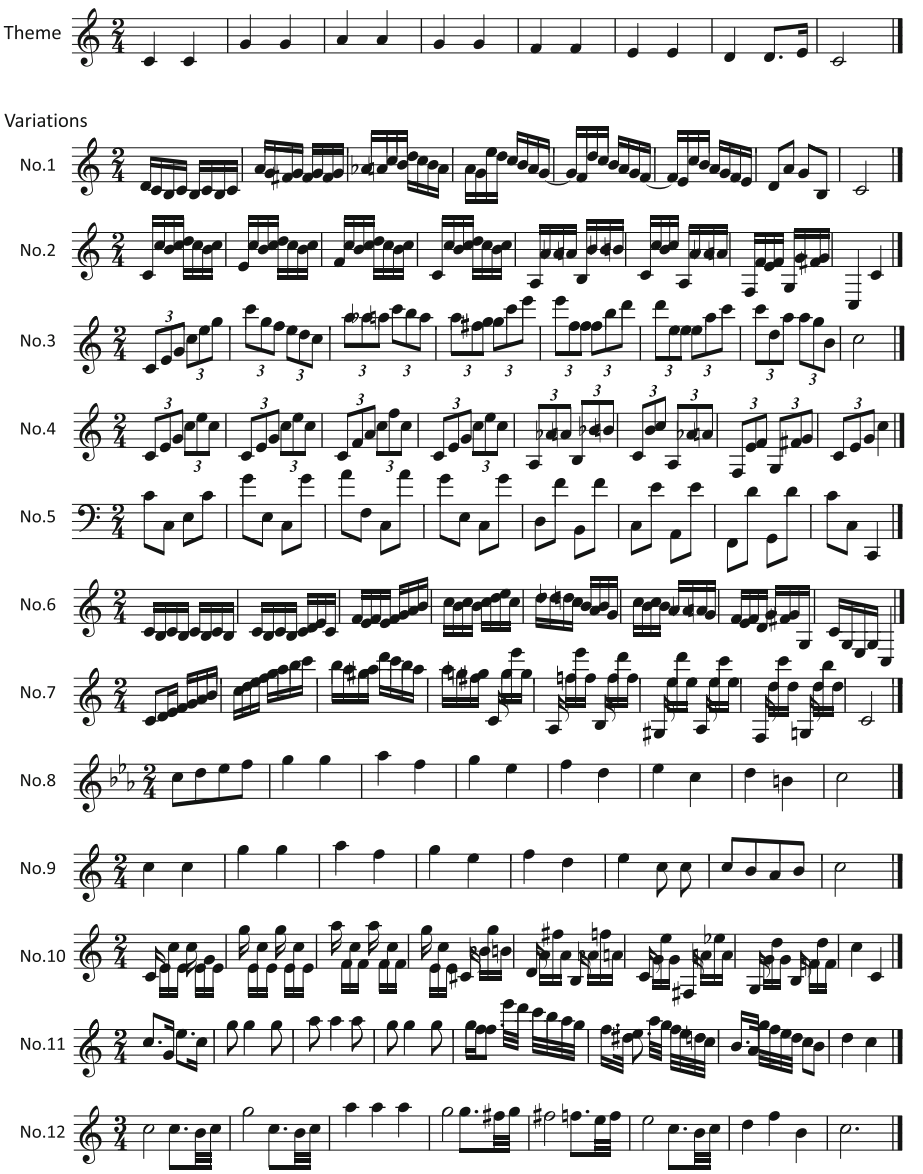



**Fig. 9.** Monophonic melodies arranged for experiment

only the upper triangle is shown. In (b), if an examinee, for instance, listen to Theme and variation No. 1 in this order, the ranking made by an examinee is found at the first row, the second column cell (−0.73). The values in (b) are the averages over all the examinees.

**Table 1.** Calculation by meet operation and psychological resemblance

(a) Theoretical estimation

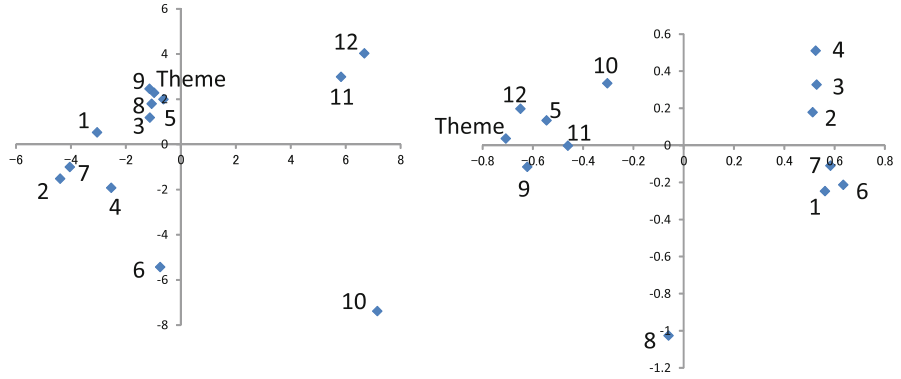
| | No.1 | No.2 | No.3 | No.4 | No.5 | No.6 | No.7 | No.8 | No.9 | No.10 | No.11 | No.12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Theme | 183 | 177 | 195 | 183 | 117 | 249 | 162 | 15 | 21 | 363 | 262.5 | 246 |
| No.1 | – | 228 | 332 | 326 | 264 | 360 | 219 | 174 | 204 | 456 | 409.5 | 421 |
| No.2 | – | – | 264 | 216 | 246 | 282 | 105 | 168 | 186 | 438 | 391.5 | 423 |
| No.3 | – | – | – | 252 | 262 | 320 | 259 | 188 | 198 | 462 | 334.5 | 379 |
| No.4 | – | – | – | – | 238 | 246 | 213 | 176 | 186 | 424 | 387.5 | 399 |
| No.5 | – | – | – | – | – | 276 | 243 | 114 | 108 | 414 | 298.5 | 325 |
| No.6 | – | – | – | – | – | – | 291 | 234 | 264 | 378 | 409.5 | 449 |
| No.7 | – | – | – | – | – | – | – | 153 | 171 | 429 | 376.5 | 400 |
| No.8 | – | – | – | – | – | – | – | – | 30 | 348 | 259.4 | 255 |
| No.9 | – | – | – | – | – | – | – | – | – | 378 | 277.5 | 261 |
| No.10 | – | – | – | – | – | – | – | – | – | – | 406.5 | 403 |
| No.11 | – | – | – | – | – | – | – | – | – | – | – | 298.5 |

(b) Rankings by human listeners (listening in row→column order)

| | Theme | No.1 | No.2 | No.3 | No.4 | No.5 | No.6 | No.7 | No.8 | No.9 | No.10 | No.11 | No.12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Theme | – | -0.73 | -0.91 | -1.09 | -0.82 | 1.18 | -1.00 | -1.45 | -0.64 | 1.36 | 0.64 | 0.73 | 1.00 |
| No.1 | -1.00 | – | -0.82 | -0.73 | -0.91 | -0.64 | 0.36 | -0.64 | -1.45 | -0.82 | -0.82 | -1.00 | -0.64 |
| No.2 | -0.91 | -0.36 | – | -0.64 | -0.27 | -0.82 | -0.45 | -0.55 | -1.55 | -0.91 | -0.09 | -0.64 | -0.91 |
| No.3 | -0.82 | -0.45 | -0.82 | – | 0 | -0.91 | -1.00 | -0.36 | -1.36 | -0.73 | -0.64 | -0.73 | -0.91 |
| No.4 | -1.00 | -0.82 | -0.73 | 0.18 | – | -0.73 | -0.82 | -0.82 | -1.73 | -0.91 | -0.45 | -1.27 | -1.00 |
| No.5 | 1.27 | -1.18 | -0.91 | -0.91 | -0.64 | – | -0.82 | -1.09 | -1.00 | 0.73 | 0.55 | 0.36 | 0.73 |
| No.6 | -1.18 | 0.27 | -0.27 | -0.45 | -0.82 | -0.64 | – | -0.36 | -1.64 | -0.91 | -0.55 | -0.64 | -0.91 |
| No.7 | -1.18 | -0.64 | -0.45 | -0.18 | -0.82 | -0.73 | -0.64 | – | -1.18 | -0.73 | -0.36 | -0.64 | -0.73 |
| No.8 | -0.73 | -1.27 | -1.36 | -1.55 | -1.27 | -0.73 | -1.00 | -1.36 | – | -0.09 | -1.09 | -0.64 | -0.91 |
| No.9 | 1.27 | -0.91 | -0.91 | -0.73 | -1.09 | 0.91 | -1.27 | -0.82 | -0.18 | – | 0.55 | 0.45 | 1.00 |
| No.10 | 0.55 | -0.82 | -0.27 | -0.64 | -0.36 | 0.73 | -0.45 | -0.82 | -1.00 | 0.73 | – | 0.18 | 0.45 |
| No.11 | 0.64 | -0.82 | -0.91 | -0.73 | -0.91 | 0.55 | -0.91 | -1.09 | -0.73 | 0.64 | 0.27 | – | 1.00 |
| No.12 | 1.09 | -1.18 | -1.09 | -1.00 | -1.00 | 0.91 | -1.00 | -1.18 | -0.91 | 1.09 | 0.36 | 0.82 | – |

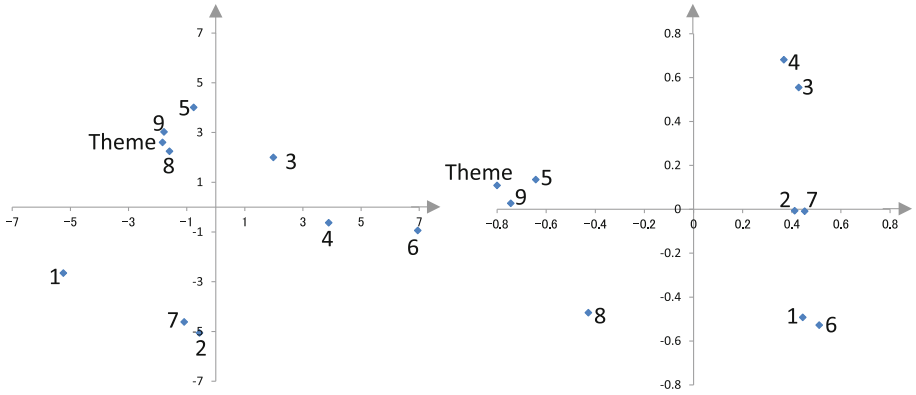
It is difficult to examine the correspondence between the results of calculated by the meet operation (a) and the psychological resemblance by examinees (b) in this distance-matrix. Then, we employ multidimensional scaling (MDS) [20] to visualize the correspondence. MDS takes a distance matrix containing dissimilarity values or distances among items, identifies the axes to discriminate items most prominently, and plots items on the coordinate system of such axes [20]. Putting it simply, the more similar items are, the closer they are plotted on a coordinate plane.

First, we use the Torgerson scaling of MDS to plot the proximity among the 13 melodies (Fig. 10), however, it is still difficult to find a clear correspondence. Therefore, we restrict plotting melodies to the theme and variations No. 1–9 (Fig. 11). Theme and No. $i$ in the figure correspond to those in Fig. 9, respectively ($i = 1..9$). The contributions in MDS are as follows: (a) Theoretical estimation: first axis (horizontal) = 0.23, second = 0.21; (b) Human listeners: first axis (horizontal) = 0.33, second = 0.17.




(a) Theoretical estimation (b) Human listener

**Fig. 10.** Relative distances among theme and all variations in mutidimensional scaling




(a) Theoretical estimation (b) Human listener

**Fig. 11.** Relative distances among theme and restricted melodies in MDS

In the figure, we can find an interesting correspondence between (a) and (b) in terms of positional relationships among 10 melodies. In both (a) and (b), we find that Theme, No. 5, No. 8, and No. 9 make a cluster; so No. 3 and No. 4 do; so No. 2 and No. 7 do. The positional relationship among the cluster of Theme, No. 5, No. 8 and No. 9, that of No. 2 and No. 7, and that of No. 3 and No. 4 resembles each other. The positional relationship between No. 1 and the others, except for No. 6, resembles, too. Since the contributions in the first axis of (a) are considered close to the second, by rotating the axes of (a) $-90°$ (counter clockwise), a more intuitive correspondence may be obtained. On the other hand, the discrepancy between (a) and (b) is seen, too; the positional relationship between No. 6 and the others is significantly different. From the above, we argue that the operations on time-span trees of our framework are viable to a certain extent.

At present, our framework disregards matching of pitch events; the meet/join operations take only the information of the configuration of time-spans. Even without matching of pitch events, however, the result of meet/join operation more or less reflects the relationship among pitches indirectly. This intuition is justified as follows. Firstly, the relationship among pitches within a music piece influences the configuration of a time-span tree construction; in fact, some of preference rules for a time-span tree are involved with the pitch and/or duration of a pitch event. Secondly, since all melodies listed in Fig. 9 are a theme and the variations derived from it, we can assume that the chord progressions of them are basically the same with each other. Thus, join/meet operations do not need to take into account the difference between chords of the two inputs.

In the sense of the second justification, it is noteworthy that Theme (or No. 5, No. 9) and No. 8 are positioned close to each other in Fig. 11 (a) Theoretical estimation, while they are not in (b) Human listener. Because No. 8 has the almost the same rhythmic structure as Theme but is in the C minor key, the time-span trees are similar while the superficial impression to human listeners are not. On the other hand, the surface structures of the No. 1 and No. 6 are almost made of 16th notes and the both are chromatic (Fig. 9). Thus, the impressions of No. 1 and No. 6 to human listeners are similar as in Fig. 11 (b), and they are positioned close to each other. However, these two variations have fairly different chord progressions; No. 1 = I→V→IV→I→IV→I→II V→I, and No. 6 = I→I→IV→I→V→I→II V→I. Accordingly, the time-span trees of No. 1 and No. 6 become distinctive in theoretical estimation, and thus they locates at the two extremes in Fig. 11(a).

## 7 Concluding Remarks

We assumed that cognitive similarity should reside in the similarity of time-span trees, that is, the reduction ordering in time-span trees were heard similarly in the order of resemblance to human ears. Based on this assumption, we proposed a framework for representing time-span trees and processing them in the algebraic manner. In this paper, we examined the validity of the framework through the experiments to investigate the correspondence of theoretical similarity with psychological similarity. The experimental results supported the convincing correspondence to some extent.

Here we have three open problems. Firstly, we exclude variations No. 10–12 for visualization in Fig. 11. Here, we need to consider why variations No. 10–12 could not achieve a higher correspondence. As possible reasons, we speculate No. 10 contains reharmonization, No. 11 features ornamental 32nd notes and No. 12 is a music piece in triple time. Thus, we are interested in an even more generalized distance that is robust to these variations.

Secondly, in terms of the contributions in MDS, the third axis in theoretical estimation which is not depicted in Fig. 11 is 0.17, and that in human listeners 0.16. Since the third axis is still relatively significant, the dimension may reveal another hidden grouping among 13 music pieces.

Thirdly, in several cases, the join operation could not be calculated as was pointed out in Fig. 6. Besides, polyphonic melodies could not be handled by our framework. Since these limitations degrade the applicability of our framework, we should extend the representation method of a music piece and tolerate the condition of the join operation, preserving the consistency with the meet operation and vice versa.

Furthermore, future work includes building a large corpus that contains more diverse melodies and conducting experiments for us to investigate the correspondence of theoretical similarity with psychological similarity in more detail. In addition, to achieve the similarity that truly coincides with our cognition, we should formalize the operations on pitch events.

**Acknowledgments.** This work was supported by JSPS KAKENHI Grant Numbers 23500145 and 25330434. The authors would like to thank the anonymous reviewers of CMMR2013 for their constructive comments to improve the quality of the paper.

## References

1. Bod, R.: A unified model of structural organization in language and music. J. Artif. Intell. Res. **17**, 289–308 (2002)
2. Carpenter, B.: The Logic of Typed Feature Structures. Cambridge University Press, Cambridge (1992)
3. de Haas, W.B.: Music information retrieval based on tonal harmony. Ph.D. thesis, Utrecht University (2012)
4. ESCOM: Discussion Forum 4A. Similarity Perception in Listening to Music, MusicæScientiæ (2007)
5. ESCOM: 2009 Discussion Forum 4B. Musical Similarity. Musicae Scientiae
6. Forte, A., Gilbert, S.E.: Introduction to Schenkerian Analysis. Norton, New York (1982)
7. Grachten, M., Arcos, J.-L., de Mantaras, R.L.: Melody retrieval using the Implication/Realization model. 2005 MIREX. http://www.music-ir.org/evaluation/mirexresults/articles/similarity/grachten.pdf. Accessed 25 June 2013
8. Hamanaka, M., Hirata, K., Tojo, S.: Implementing "A Generative Theory of Tonal Music". J. New Music Res. **35**(4), 249–277 (2007)
9. Hewlett, W.B., Selfridge-Field, E. (eds.): Melodic Similarity - Concepts, Procedures, and Applications. Computing in Musicology, vol. 11. The MIT Press, Cambridge (1998)
10. Lerdahl, F., Jackendoff, R.: A Generative Theory of Tonal Music. The MIT Press, Cambridge (1983)
11. Marsden, A.: Generative structural representation of tonal music. J. New Music Res. **34**(4), 409–428 (2005)
12. Mirex, HOME. http://www.music-ir.org/mirex/wiki/MIREX_HOME. Accessed 25 June (2013)
13. Pampalk, E.: Computational models of music similarity and their application in music information retrieval. Ph.D. thesis, Vienna University of Technology, March 2006
14. Pearce, M.T., Wiggins, G.A.: Expectation in melody: the influence of context and learning. Music Percept. **23**(5), 377–405 (2006)

15. Rizo-Valero, D.: Symbolic music comparison with tree data structure. Ph.D. thesis, Departamento de Lenguajes y Sistemas Informatícos, Universitat d' Alacant (2010)
16. Schedl, M., Knees, P., Böck, S.: Investigating the similarity space of music artists on the micro-blogosphere. In: Proceedings of ISMIR 2011, pp. 323–328 (2011)
17. Tojo, S., Hirata, K.: Structural similarity based on time-span tree. In: 10th International Symposium on Computer Music Multidisciplinary Research (CMMR), pp. 645–660 (2012)
18. van Kranenburg, P.: A computational approach to content-based retrieval of folk song melodies. Ph.D. thesis. Utrecht University (2010)
19. Volk, A., Wiering, F.: Tutorial musicology, Part 3: music similarity. In: 12th International Society for Music Information Retreival Conference, ISMIR 2011 (2011). http://ismir2011.ismir.net/tutorials/ISMIR2011-Tutorial-Musicology.pdf. Accessed 15 June 2013
20. Wikipedia, Multidimensional scaling. http://en.wikipedia.org/wiki/Multidimensional_scaling. Accessed 15 June 2013
21. Wikipedia, Temporal Logic. http://en.wikipedia.org/wiki/Temporal_logic. Accessed 25 June 2013