No Longer 'Somewhat Arbitrary': Calculating Salience in GTTM-Style Reduction

Alan Marsden
Lancaster University, UK
a.marsden@lancaster.ac.uk

Satoshi Tojo
JAIST, Japan
tojo@jaist.ac.jp

Keiji Hirata
Future University Hakodate, Japan
hirata@fun.ac.jp

ABSTRACT

Following earlier work on the formalisation of Lerdahl and Jackendoff's Generative Theory of Tonal Music (GTTM), we present a measure of the salience of events in a reduction tree, based on calculations relating the duration of time-spans to the structure of the tree. This allows for the proper graphical rendition of a tree on the basis of its time-spans and topology alone. It also has the potential to contribute to the development of sophisticated digital library systems able to operate on music in a musically intelligent manner. We present results of an empirical study of branch heights in the figures in GTTM which shows that salience calculated according to our proposals correlates better with branch height than alternatives. We also discuss the possible musical significance of this measure of salience. Finally we compare some results using salience in the calculation of melodic similarity on the basis of reduction trees to earlier results using time-span. While the correlation between these measures and human ratings of the similarity of the melodies is poor, using salience shows a definite improvement. Overall, the results suggest that the proposed definition of salience gives a potentially useful measure of an event's importance in a musical structure.

CCS CONCEPTS

• **Applied computing → Sound and music computing**;

KEYWORDS

GTTM, Tonal reduction, Salience

1 INTRODUCTION

Several theories of music describe musical structures in a hierarchical manner, most notably the *Generative Theory of Tonal Music* (GTTM) [9] and the theory of Heinrich Schenker [15]. Some notes in a piece of music have a greater structural importance than others.

Different versions of a folk song or hymn tune often have slight differences among the less important notes. Jazz improvisations can introduce new decorative chords or simplify a harmonic sequence by leaving out inessential chords. A variation of a melody in Classical music similarly preserves the basic structure while presenting a different sequence of notes. This presents two kinds of challenge for digital library systems for musicology. Firstly, systems will sometimes contain or use information about the hierarchical structure of pieces of music, and the system needs to be able to present this information to the user in a meaningful and effective manner. Secondly, in order to perform operations such as finding variations of a theme, a system will need to have or calculate information about the structural importance of events in pieces of music.

Lerdahl and Jackendoff present heirarchies in trees. The branching of the tree corresponds to the introduction of less essential, more decorative, notes in the musical structure. The musical process of reduction, progressively removing decorative elements to reveal the essential musical structure, corresponds to progressively pruning leaves from the tree. An example is shown in Figure 1 [9, Fig. 5.8, p. 115]. Schenker, by contrast, presents hierarchies in *levels* which resemble standard music notation, where the higher levels contain only the more structurally important notes. Lerdahl and Jackendoff use a presentation like this also, calling it a 'secondary notation' [9, p. 117]. An extract of a corresponding example is shown in Figure 2.

The two kinds of notation are clearly related. Derivation of a tree structure from a Schenkerian analysis presenting levels is discussed in [11, p. 416]. In the opposite direction, ideally one can derive a level of reduction from a tree representation simply by cutting off all the branches below an appropriately positioned horizontal line and writing down the corresponding pitch events which remain, as illustrated in Figure 1. However, trees are not always so simple, and Lerdahl and Jackendoff confess that in their analysis of the opening of Mozart's G-minor symphony "the assignment of levels in the time-span reduction [is] somewhat arbitrary." [9, p. 258] In this paper we propose a way to determine the salience of a branch (corresponding to its structural importance)[1] based on the hierarchical structure of the tree and the durations of the pitch events in the music.[2]

---

[1] Our use of the term 'salience' is different from that of Lerdahl and Jackendoff. They use the term 'surface salience' and contrast it with 'structural importance' [9, p. 108–109]. In his later book, Lerdahl uses the terms 'salience' and 'stability' to refer to factors which contribute to, but do not of themselves determine, structural importance [10, p. 313–321]. We prefer to use 'salience' as a synonym for 'structural importance' so as to have a single-word term for this important concept. The phrases 'surface salience' and 'psychoacoustic salience' could be used for the factor Lerdahl refers to. In summary, in our usage 'salient' means 'important' rather than 'sticking out'.

[2] We do not take pitch information into account beyond its important influence on the tree structure. Whether or not salience can be adequately calculated without specifically taking pitch information into account is a question for future research.
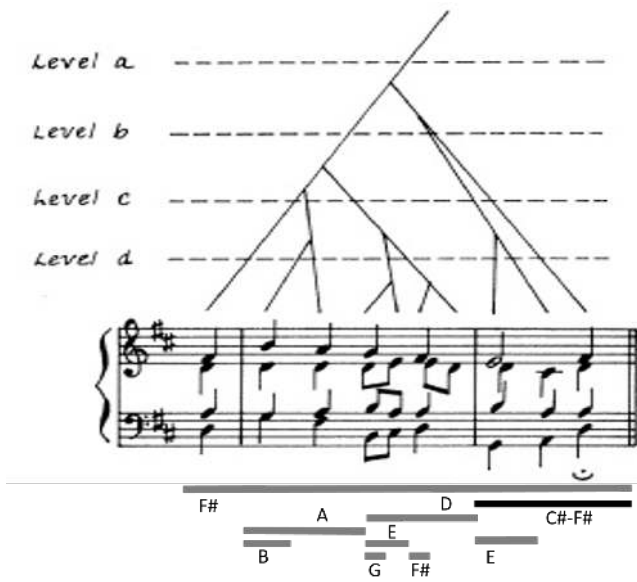
Figure 1: Tree in GTTM with maximum time-spans, extracted from [9, Fig. 5.8, p. 115]. The horizontal lines we have added at the bottom of the figure show the maximum time-spans. See Section 2 for further explanation.

Figure 2: Partial secondary notation corresponding to Figure 1, extracted from [9, Fig. 5.8, p. 115]. We have reversed the order of levels b and c in order to conform to the vertical order of Figure 1 and the normal order in Schenkerian analyses.

This has potential applications in the field of digital libraries for musicology by allowing reduction trees to be displayed algorithmically with branch heights corresponding to the salience of the events in the music. A digital symbolic representation of a reduction tree can therefore be given an appropriate graphical form. An analysis tree could be graphically rendered on the fly from a representation which gave just the structure of the tree and the durations of the notes rather than requiring storage of an image of the tree.

We hope that in the longer term our quantification of salience can lead to more effective systems to derive structural analyses from musical surfaces. This will in turn lead to more effective tools for digital libraries for musicology where, as mentioned above, it is necessary to deal with musical data in a musically intelligent fashion (*e.g.*, for determining similarity, recognising variations, finding patterns or making summarisations). It could also contribute to other applications such as generation of music through morphing and elaboration.
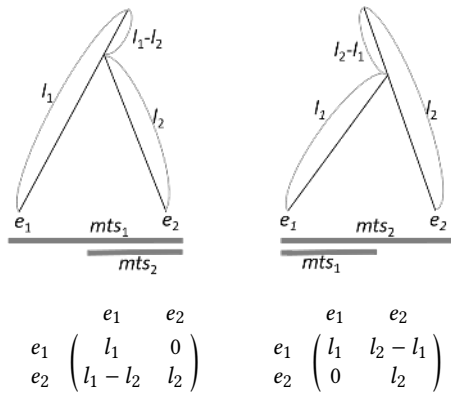
In this paper we first discuss the concepts of reduction, tree structure and salience (Section 2) and explain our motivation. Then we present our essential proposal for quantifying salience in a hierarchical structure represented by a binary tree (Section 3). This is followed by a discussion of evident and possible principles concerning branching height and the shape of trees (Section 4), and we present evidence that the branch height used in the printed illustrations in GTTM correlates well with our measure of salience (Section 5). In Section 6 we discuss possible music-theoretic bases for this measure of salience, drawing particularly on the idea of musical information. Finally, as a first test of whether salience as defined here has wider application than simply in the drawing of reduction trees, we present some evidence that salience can improve the modelling of judgements of melodic similarity (Section 7) .

2 TREE STRUCTURE AND SALIENCE

GTTM has been widely influential, but it is not universally accepted as a valid and reliable theory of musical cognition. (For a recent survey see [5] and the special issue of *Musicae Scientiae* dedicated to a symposium on GTTM (vol. 14, issue 1_suppl, 2010). For discussion of its cognitive validity see [8].) While the research reported here is based on this specific theory, it is not completely dependent on it. The findings apply to any theory which presents musical structure in terms of binary trees. The fundamental idea of hierarchical music structure is widely accepted, and supported by empirical evidence drawn from sources ranging from human's association of a musical extract with its reduction [1, 13], to Classical variations [12], to memory and performance errors [17].

The contrast mentioned above between tree notations and levels (Lerdahl and Jackendoff's 'secondary notation') can be re-expressed as a question of whether salience is a total order (as implied by levels) or only a partial order (as implied by trees). In other words, given two note events a and b, where neither is a descendent of the other (*i.e.*, there is no exclusively downward or upward path in the tree from one to the other), can their relative salience be determined? A representation of musical structure in levels shows relative salience in such circumstances, but a representation in a tree does not, unless, as Lerdahl and Jackendoff's diagrams imply, the relative vertical positions of the branching points on the page carry some meaning.

Our interest in the topic originates in our use of matrices to represent tree structures [19]. The connections in a tree can be represented in a matrix in which each row and column corresponds to a pitch event in the music. A value greater than 0 in cell (i, j) indicates that the event i is connected and subordinate to event j. The structure of a tree can therefore be represented in a square matrix with values 0 or 1 in cells as appropriate, and calculations

$$\begin{array}{cc} & \begin{array}{cc} e_1 & e_2 \end{array} \\ \begin{array}{c} e_1 \\ e_2 \end{array} & \begin{pmatrix} l_1 & 0 \\ l_1 - l_2 & l_2 \end{pmatrix} \end{array} \qquad \begin{array}{cc} & \begin{array}{cc} e_1 & e_2 \end{array} \\ \begin{array}{c} e_1 \\ e_2 \end{array} & \begin{pmatrix} l_1 & l_2 - l_1 \\ 0 & l_2 \end{pmatrix} \end{array}$$

Figure 3: Matrix representation with relation between MTS and branch height [19, fig.4].

on the tree (such as determining whether or not an event a is on the reduction path for event b, the 'reachability' relation between two events) can be performed through algebraic operations on the matrix.

In earlier work, the concept of **maximum time-span (MTS)** has been shown to be useful [6, 18]. If a pitch event does not have branching, *i.e.*, there is no more subordinate pitch event and it is a leaf of the tree, its MTS is the original pitch length. At the other extreme, the MTS of the event that reaches the top of the tree is the whole length of the music piece. We show MTSs in the bottom of Figure 1 by thick gray lines. The conventions are: (1) the length of each pitch represents maximum time-span, (2) the black line represents a cadential retention, *i.e.*, two pitch events are paired and treated as one event, (3) a lower span is subordinate to its direct higher span, (4) one of the two ends of each span must coincide with an end of some upper spans, and (5) the spans on the same horizontal level imply the same salience.

We speculate that a more useful matrix representation is possible using not just the values 0 and 1, but values related to the MTS for each event. In [19] we demonstrated how placing the MTS values on the diagonal of the matrix allows matrix multiplication to be used (as in matrices using only 0 and 1) to determine reachability without destroying the MTS information in the matrix. What value should be used off the diagonal to indicate connection between events? This value should give some information about the relationship between the events, just as the height of branching in a tree diagram gives information about the relative structural importance of events, by indicating the 'level' of an event. In [19] we proposed to use the difference between the MTS of the superordinate event and the MTS of the subordinate event. This allowed a simple relationship between MTS and branch height (Figure 3).

However, keeping that simple relationship will lead, in some circumstances, to tree diagrams which violate the basic principle that structurally more important events branch from the stem at a higher point than less important events. In Fig. 4 we show how a structurally more important event (the last note in the second example) would branch from the tree at a lower point than a less important event (the second note in that example).
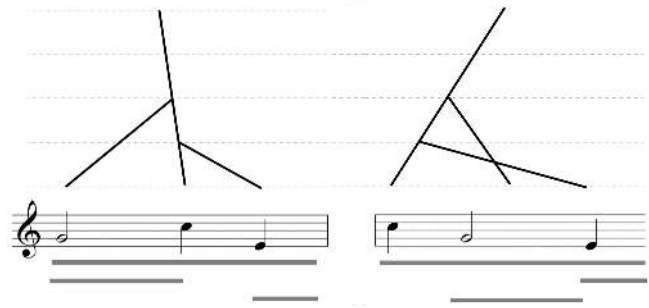

Figure 4: Well formed tree and badly formed tree with crossing branches.

3 QUANTIFYING SALIENCE

Before introducing our solution to this problem, we explain the terms we will use. A branching point in a time-span tree structure corresponds to a point in the corresponding process of reduction when one of two events which make up a time-span is selected as subordinate to the other event, leaving that event as the 'head' of the new combined time-span. Alternatively, we can view the tree as representing a process of generation rather than reduction, and the branching point corresponds to a point in this generative process when a single longer time-span is split into two time-spans, introducing a new pitch event which ornaments the superordinate event without replacing its structural role, just as a trill or grace note in performance ornaments the principal written note.

There are thus three time-spans involved at every branching point, a long one corresponding to the 'parent' and two shorter ones corresponding to the 'children'. The relationship of these three time-spans is illustrated in Figure 5, in which we draw the time-spans across the corresponding branches instead of below them as we have in other figures. The time-span of the subordinate child is the maximum time-span (MTS) of the event corresponding to the new branch, the longest time-span in which that event is the head (most salient). The duration of this time-span is the sum of the durations of all the surface events which make up this time-span. The duration of the time-span of the superordinate child is similarly the sum of the durations of all the surface events which make up that time-span, but this is not that event's maximum time-span. (Because this is the superordinate child, the event corresponding to this child is also the head of the time-span corresponding to this branching point, the parent time-span (see below), which must be longer in duration.) We call this instead the **intermediate time-span (ITS)**, which is the longest time-span of which the event is head *up to that point in the tree*. The three time-spans involved in a branching point are therefore the MTS of the subordinate child, the ITS of the superordinate child, and the time-span of the parent, which we will call the **parent time-span (PTS)**, which is the union of the MTS and ITS of the two children.

At the branching point above this one, the PTS becomes either the MTS (if this event is the subordinate child, or indeed if there is no branching point above) or the ITS at that branching point (if this event is the superordinate child). An event (corresponding to a branch or stem of the tree) thus has one MTS, and zero or more

$$PTS = ITS + MTS$$
$$salience(e_2) = \max(ITS, MTS)$$
$$PTS\ /\ 2 \le salience(e_2) < PTS$$
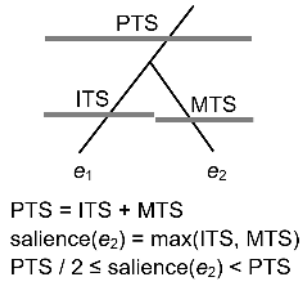
Figure 5: Illustration of the relationship of maximum time-span (MTS), intermediate time-span (ITS) and parent time-span (PTS) at a branching point. The salienceof a branch is the maximum of the ITS and MTS at the point where the branch joins to the tree.

ITSs. To know the ITS to be used in a calculation, we need to specify the branching point concerned as well as the event in question.

As discussed above, branch height ideally corresponds to salience or structural importance. We propose to define the **salience** of an event as the duration of the *maximum* of the time-spans of the two children at the branching point when the event is generated, or where it is reduced, *i.e.*, the greater of the MTS of the subordinate child (the event on the branch) and the ITS of the superordinate child (the event on the stem) at that branching point. (In the case of an event which is never reduced, *i.e.*, the root of the tree, the salience is defined as equal to the event's MTS.) The salience of an event is thus always greater than the salience of every event which occurs below the branching point where that event connects to the tree structure. We can therefore use salience as a measure of branching height and be sure that trees malformed in the manner of Figure 4 will never occur.

We emphasise that the salience of an event, by this definition, does not depend on the duration of the event alone. The event's duration places a lower bound on the salience (it can never be less than the event's duration) but the salience can depend also on the duration of its sibling event. Thus while we do not directly measure other factors such as pitch, to the extent that these factors influence the branching structure of the tree, they also influence the salience of events.

4 BRANCH HEIGHT

In this section we examine salience as defined above in relation to the height of branches in the display of a tree. By 'height' we mean here the vertical distance on the page between the leaves of the tree and the point where the new branch branches off the tree. This is different from the use of the term 'height' in classical graph theory, where height is calculated by counting what we here call branching points and so depends only on the structure of a tree, not on how it is displayed. Below we will use the term 'depth' in the classical manner. In summary, according to our usage here, height depends on how a tree is displayed, depth depends only on the structure of the tree.

Lerdahl and Jackendoff do not specify any details for how a tree should be drawn but we can infer certain principles from the examples in their book.

- All branches are straight lines.
- There are no crossing branches.
- The leaves of the tree (the ends of the lines) are positioned on an invisible horizontal line at the bottom.
- The leaves are ordered along that line according to their temporal position in the music represented.
- The root of the tree is at the top, positioned approximately centrally in relation to the leaves.

Although these principles constrain the height of each branch relative to the total height of the tree, they are not sufficient to determine that height. Certain ways of calculating branch height are guaranteed to ensure that there are no crossing branches, provided the other principles (or, to be precise, all except the last) are also followed. We have tested three candidate methods which have this property, listed below. In each of these *hmax* is the height of the highest branching point in the tree and *MTSmax* is the duration of the MTS of the head of the tree (*i.e.*, the duration of the entire piece of music analysed). The objective is to determine *hbranch*, the height of the point where the branch in question connects to the tree.

(1) **Depth.** Let the **depth** of a branching point be the number of branching points above it in the tree. The height of a branching point could be related linearly to its depth. Let *dbranch* be the depth of the branching point in question and *dmax* be the greatest depth of any branching point in the tree. By this principle the height of each branch is calculated as

$$hbranch = \frac{hmax \cdot (dmax - dbranch + 1)}{dmax + 1}$$

Since, by definition, the depth of any branching point is less than the depth of all the branching points below it, there can be no crossing branches.

(2) **PTS.** This principle relates branch height to the duration of the parent time-span at the branching point (*PTSbranch*). By this principle the height of each branch is calculated as

$$hbranch = \frac{hmax \cdot PTSbranch}{MTSmax}$$

By definition, *PTSbranch* must be greater than the ITS and MTS at this point (and at all branching points below) and since the value of PTS at the next branching points below (if any) is equal to these ITS and/or MTS values, all lower branches will connect at a lower height, and crossing branches are avoided.

(3) **Salience.** This principle relates branch height to salience. Let *sbranch* be the salience at a branching point and *smax* be the salience at the highest branching point in the tree (i.e., the salience of the highest branch which is not the root of the tree). The value of *smax* is constrained by *MTSmax* according to the inequality between PTS and salience shown in Figure 5. By this principle the height of each branch is calculated as

$$hbranch = \frac{hmax \cdot sbranch}{smax}$$
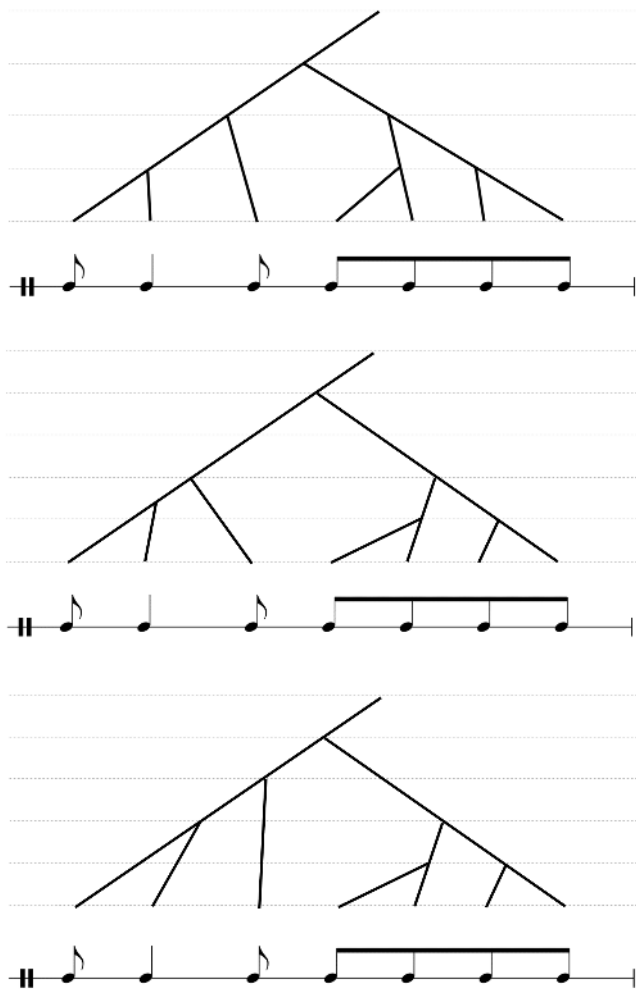
Figure 6: Examples of trees drawn according to the three candidate height principles. Top: Depth; Middle: PTS; Bottom: Salience.

> Since salience is the maximum of MTS and ITS at each point, is must be greater than the salience at all branching points below, and crossing branches are once again avoided.

Examples of trees drawn according to these three principles are shown in Fig, 6.

Determining branch height by branching depth, the first candidate above, results in trees where levels can be separated by equal vertical distances. The tree from GTTM shown in Figure 1 shows approximately this property.

One disadvantage of trees shaped like this is that subtrees become progressively more skewed the closer they are to the beginning and end of the music. In the tree in Figure 1, for example, the subtrees around levels b and c representing the first three beats, middle two beats, and last three beats, respectively, all have a highest point which is off to one edge of the corresponding time-span or, in the case of the last of these three subtrees, even beyond the edge of the

time-span. Under this method of calculating branch height, subtrees are not similar in their basic shape to the full tree.

Determining branch height by PTS, however, causes the height of each branching point to be related to the duration of the time-span covered by that point, and so all trees and subtrees are, on average, formed of similarly shaped triangles. A disadvantage, though, is that levels are now compressed towards the bottom of the tree and the top of the tree diagram consists mostly of blank space.

Since salience is constrained by PTS (see Figure 5), trees drawn according to the third formula above have similar basic properties to those described above for trees where branch height is determined by PTS.

5 BRANCH HEIGHTS IN GTTM

To test the three proposals described in Section 4, the branch heights of the 28 trees in GTTM where the relative duration of events could be determined were measured. To do this, we placed a ruler along the bottom of each tree (which revealed that in a few cases the leaves are not exactly on a horizontal line, but the deviation is never more than 2mm) and used a second ruler held at right angles to this one to measure the height of each branching point to the nearest millimetre. This height, the duration of the surface time-span at the end of the branch and the direction of the branching (left or right) were recorded. In cases where a tree did not give a complete analysis for a piece, because the lower levels were not shown, a judgement was made about the most musically appropriate time-span to measure, taking into account other information available such as any notation showing grouping and metre. Specially written software read this record, constructed each tree, and calculated the lengths of time-spans and saliences at each branching point.[3]

To facilitate analysis, branch heights were normalised by dividing each height by the maximum height in the corresponding tree. Time-span durations and saliences were normalised by dividing by the total duration spanned by the corresponding tree (the MTS of the root of the tree).

Table 1 shows the correlations calculated from the resulting 394 data points. Assuming that the diagrams in GTTM are drawn from a population of all possible reduction-tree diagrams, the variance in the correlation of actual branch height with calculated branch height for each method can be estimated using jackknife resampling [2]. The significance of the differences shown in Table 1 was calculated using a two-tailed Student T-test. All differences were found to be significant with $p < 0.001$. We can therefore conclude that PTS and salience are correlated more strongly with branch height in the trees in GTTM than is depth. Of these two, salience has the slightly stronger correlation suggesting it is the best candidate to use when calculating the branch height to use when drawing trees.[4]

---

[3]All the data and software referred to in this paper can be found at http://www.lancaster.ac.uk/people/marsdena/research/salience.

[4]An examination of the correlations within each tree diagram rather than across the dataset as a whole presents a somewhat different picture. PTS and Salience once again clearly outperform Depth ($p < 0.01$), but they do not significantly differ from each other when correlations are compared in this way. Within each tree, there are likely to be other factors affecting branch height which should be examined in a deeper study of the graphic display of reduction trees.

Table 1: Correlation with Branch Height

Measure	Coefficient of correlation
Depth	0.826
PTS	0.912
Salience	0.926
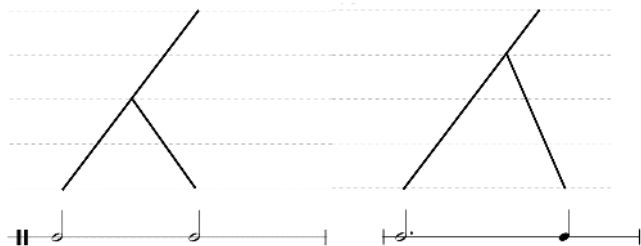
Figure 7: Equal and unequal divisions of a time-span.

6　SALIENCE IN MUSICAL INFORMATION

In this section and the following one, we consider whether salience is not only useful for drawing trees but also has a deeper musical meaning.

The issue we consider here is the salience of equal and unequal divisions of a time-span. The salience of an unequal division of a time-span is, according to our definition, always greater than the salience of an equal division of the same time-span, because the maximum of the two child time-spans will always be greater in an unequal division than in an equal division. For example, in the trees shown in Figure 7 the salience of the second note in the second tree is higher than the salience of the second note in the first tree.

It seems counter-intuitive that a short note should be more salient than a long note. However, it is reasonable that the salience of an event should depend not just on the properties of the event itself but also on the context in which that event is placed. A single dot on an otherwise blank piece of paper is much more salient than a similar sized dot on a piece of paper with lots of other marks.

According to Shannon's information theory, the quantity of information in an event is inversely related to its likelihood [16]. If a time-span is divided, the most likely division is into equal parts, so an unequal division contains more information than an equal one. To specify an equal division only requires stating that a time-span is divided. To specify an unequal division also requires specification of the proportion into which it is divided.

Most music is characterised by a *metre* such as 4/4 or 6/8. Normally this is explained as a pattern of beats but in practice not every beat corresponds to an actual event in the music. A more consistent definition of metre is that it sets a framework of probabilities for when events are likely to occur in the music [3]. Following the logic of the paragraph above, events which do not fall on the main beats are less likely, and so carry more information, than events which fall on the main beats. In the case of simple binary metres like 4/4 and 2/4, this means that salience and information content (at least in terms of timing and metre) are consistent in simple two-note configurations, provided the first note falls on a strong beat than

the second, as is most common. In the case of ternary metres such as 3/4, however, salience and information content diverge. In these cases division of the time-span of a bar (measure) into two equal halves is unlikely (high information content) but still of minimum salience. This is a topic we will need to revisit in future research.

7　SALIENCE AND SIMILARITY

In this section we revisit data on melodic similarity used in an earlier study of distance measured according to the relations between reduction trees. The melodies concerned were derived from a set of variations, so it can be assumed that each melody has some degree of similarity to others in the set. Human ratings of the similarity between each pair of melodies were obtained from a panel of 11 listeners. For details, see [6, 7].

We have defined operations of *meet* (⊓) and *join* (⊔) on two reduction trees, finding, respectively, the maximum tree which is common to the two original trees and the minimum tree which is an elaboration of both original trees [18]. This operation provides a basis for the morphing of one melody into another and for measurement of the distance between two melodies on the basis of the steps required to transform the tree of one melody into that of the other by first removing branches from the first tree to arrive at the *meet* tree and then adding new branches to arrive at the second tree [7].

The distance between two trees x and y can then be measured as the quantity of information added or taken away in transforming tree x into tree y by first removing branches from x to arrive at the tree which is $x \sqcap y$ and then adding branches to that to arrive at the tree y. If $h(a)$ is the total quantity of information in a, the distance between two trees $d(x, y)$ can be calculated as

$$d(x, y) = h(x) - h(x \sqcap y) + h(y) - h(x \sqcap y)$$

In earlier work, the quantity of information was taken to be measured by the time-span of branch added or removed (more strictly the time-span of the head of the branch added or removed). In other words, h was equivalent to calculating the *total maximum time-span* (*tmts*) of a tree, the sum of the durations of the maximum time-spans of all the events in the tree [18].

One problem with this simple approach is that when comparing two trees cases arise when the time-spans of corresponding events are not equal and cases arise when the branching does not correspond (*e.g.*, one tree branches to the right where the other branches to the left). In order to allow a distance to be calculated between any two trees, the calculation of *meet* is generalised to effectively ignore branches which do not correspond and to define the time-span of events in the *meet* tree to be that time-span which is common to both trees (*i.e.*, the temporal overlap) [6].

In earlier work, this method of measuring melodic similarity was tested on the melodies from Mozart's variations of *'Ah, vous dirai-je, maman'*, K.265/300e. A panel of listeners (11 university students) also rated the similarity of the melodies, and the two sets of distance measures were compared using multidimensional scaling (MDS). Similar clusters emerged in the MDS results for the two measures, indicating that the reduction distance had some musical validity [6, 7].

Here we examine whether using salience, as defined above, rather than time-span might produce a better measure of distance based

on reduction. Following the logic of the previous section, we hypothesise that salience provides a better measure of the quantity of information added or removed when transforming tree x into tree y via $x \sqcap y$. We tested three ways of measuring this quantity of information:

(1) **Time-span overlap (TSO).** This is the original method, where the time-span of an event in the *meet* tree is the overlap of the corresponding time-spans in the original trees (possibly null) and h is the total maximum time-span.

(2) **Time-span duration (TSD).** In this method, h is once again the total maximum time-span of a tree, but the duration of a time-span in the *meet* tree is the lower of the durations of the corresponding time-spans in the original trees. In other words, how much the corresponding time-spans overlap is ignored and the shorter of the two corresponding time-spans is used in the *meet* tree.

(3) **Salience (SAL).** In this method, h is the sum of the salience of all the events in a tree, and the salience of a branch in the *meet* tree is the lower of the saliences of the corresponding branches in the original trees.

We also took the opportunity to examine the effect of a modification to the *meet* algorithm as defined in [6]. In the original definition, when the branching in two trees does not correspond (*i.e.*, one tree has right-branching where the other has left-branching), the two branches are excluded from the resulting *meet* tree. In Prolog code, the recursive part of the original algorithm is:

```
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).
```

The operators < and > are used in the Prolog representation of a tree structure for left and right branching, respectively. The first two clauses above therefore handle cases where the branching in the two trees matches, and the second two clauses handle cases where the branching does not match. We experimented here with a modification of the algorithm in the cases of non-matching branching. The modified algorithm selects the 'maximum' of two possible results,

each of which excludes the branch in just one of the trees. This effectively allows a realignment in the correspondence of time-spans in the two input trees, so we call this *meet with realignment*. The adapted algorithm replaces the last two clauses with the following:

```
meet(X,Y,Meet) :-
    X = (Xp > _),
    Y = (_ < Yp),!,
    meet(X,Yp,Meet1),
    meet(Xp,Y,Meet2),
    maxh(Meet1,Meet2,Meet).
meet(X,Y,Meet) :-
    X = (_ < Xp),
    Y = (Yp > _),!,
    meet(X,Yp,Meet1),
    meet(Xp,Y,Meet2),
    maxh(Meet1,Meet2,Meet).
```

where maxh/3 is a predicate which ensures the third argument is whichever of the first two arguments maximises the value of h (however that is defined; see above). When h is measured by the original method (called TSO above), the result is rarely different because the maximum overlap in time-spans generally arises from not realigning their correspondence. The differences in the cases of the other two methods of calculating h are generally greater.

Finally, a method of normalising distances is required, because the raw distance between two large and complex trees is potentially much greater than that between two simple trees. A method of calculating a normalised similarity proposed in earlier work [4] is as follows:

$$sim1(x,y) = 1 - \frac{h(x) - h(x \sqcap y)}{2 \cdot h(x)} - \frac{h(y) - h(x \sqcap y)}{2 \cdot h(y)}$$

We used this method of normalisation and also experimented with a second method, which gives equal weight to the 'lost' and 'added' information but is not strictly a normalisation since it could result in a value less than 0:

$$sim2(x,y) = 1 - \frac{h(x) - h(x \sqcap y) + h(y) - h(x \sqcap y)}{\max(h(x), h(y))}$$

The previous method of comparing computed distances with human similarity ratings through MDS plots does not afford a simple way of comparing the efficacy of one method of computing distances with another. We have therefore measured the correlation between the average human ratings of the similarity of each pair of melodies and the normalised similarities between those melodies, computed by the various different methods described above. The coefficients of correlation are shown in table 2. Using jackknife resampling to estimate variance, the differences between the three cataegories TSO, TSD and SAL were found to be highly significant ($p < 0.001$), but within categories the only significant differences are in SAL, where the difference between strict+sim1 and realigned+sim2 is highly significant and other differences except that between realigned+sim1 and strict+sim2 are marginally significant ($0.037 < p < 0.073$). The correlation coefficients are all so low as to indicate effectively no correlation at all, showing that there are many other factors in determining similarity ratings than reduction distance. (Recall that this method takes no account,

Table 2: Correlation with Similarity Ratings

h measure	Meet	sim1	sim2
TSO	strict	-0.109	-0.116
TSO	realigned	-0.107	-0.115
TSD	strict	-0.042	-0.041
TSD	realigned	-0.040	-0.036
SAL	strict	0.052	0.072
SAL	realigned	0.071	0.090

when determining the reduction distance, of changes in the pitch of notes when transforming one melody to the other. On the other hand, pitch is important in determining the tree structures of the two melodies, so this is not a measurement of rhythmic similarity alone.) Nevertheless, the correlations when using salience in the calculation of distance are greater and in the expected direction, especially if realignment is allowed in calculating the meet tree. The evidence is not strong, but it does suggest that salience is of potential value in calculating the quantity of information lost and added when computing reduction distance.

8 CONCLUSION

We have proposed a new definition of the *salience* of a branch in a tree structure as the maximum of the durations of the time-spans of the events which combine at the point where a branch connects to the tree. We have presented some arguments and empirical results to support our case that this is a potentially useful definition.

Our strongest evidence concerns the use of this measure of salience in determining the height of branches when drawing tree diagrams. This has a potential direct application in musical digital library systems since it gives a way to calculate an appropriate height for each branch from the essential information in the tree structure. Given a representation of a piece of music and its tree-structure analysis, for example in a format such as proposed by Rizo and Marsden as an extension to MEI [14], it would be possible for a digital library system using this measure of salience to render analysis trees on the fly without needing to store an image representation of the tree.

We propose that this measure of salience is also of potential value in determing the 'quantity of information' carried by a branch in a tree structure. A comparison of the correlations with human ratings of melodic similarity in a set of variations by Mozart with similarity computed using salience and other methods gave support to this claim, but the evidence is not strong.

The results and discussion here show three areas demanding further research. Firstly, as observed at the end of Section 6, while our definition of salience seems to correspond well to musical salience with respect to rhythm in the case of binary metres, this is not so for ternary metres. Secondly, as noted in footnote 1 and shown by the results for melodic similarity, research is required on how to account for pitch information when calculating salience. Thirdly, further research is required into reduction distance as a measure of melodic similarity.

Looking further ahead, we hypothesise that salience as defined here could be used in the process of analysis, *i.e.*, deriving a tree

structure from the sequence of notes making up the surface of a piece of music. Salience in music has been related to several other musical details, including relationships of pitch and harmony [10]. In the process of analysis, correlating salience as indicated by other details with the salience as calculated from the tree structure might be a way of selecting between possible alternative analyses.

ACKNOWLEDGMENTS

REFERENCES

[1] Nicola Dibben. 1994. The Cognitive Reality of Hierarchic Structure in Tonal and Atonal Music. *Music Perception* 12, 1 (1994), 1–25.
[2] B. Efron and C. Stein. 1981. The Jackknife Estimate of Variance. *The Annals of Statistics* 9, 3 (1981), 586–596.
[3] Mark Gotham. 2017. Hierarchy and position usage in 'mixed' metrical structures. *Journal of New Music Research* 46, 2 (2017), 103–117. https://doi.org/10.1080/09298215.2016.1253752
[4] Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. 2015. Structural Similarity Based on Time-Span Sub-Trees. In *Mathematics and Computation in Music*, Tom Collins, David Meredith, and Anja Volk (Eds.). Springer International Publishing, Cham, 187–192.
[5] Niels Chr. Hansen. 2011. The Legacy of Lerdahl & Jackendoffâ㎝s A Generative Theory of Tonal Music: Bridging a significant event in the history of music theory and recent developments in cognitive music research. *Danish Yearbook of Musicology* 38 (2011), 33–55.
[6] Keiji Hirata, Satoshi Tojo, and Masatoshi Hamanaka. 2014. Cognitive Similarity Grounded by Tree Distance from the Analysis of K.265/300e. In *Sound, Music, and Motion: 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers*, Richard Kronland-Martinet Mitsuko Aramaki, Olivier Derrien and Sølvi Ystad (Eds.). Springer, Cham, Switzerland, 589–605.
[7] Keiji Hirata, Satoshi Tojo, and Masatoshi Hamanaka. 2016. An Algebraic Approach to Time-Span Reduction. In *Computational Music Analysis*, David Meredith (Ed.). Springer, Cham, Switzerland, 251–270.
[8] Michel Imberty. 2000. The Question of Innate Competencies in Musical Communication. In *The Origins of Music*, Björn Merker Nils L. Wallin and Steven Brown (Eds.). MIT Press, Cambridge, MA, USA, 449–462.
[9] Fred Lerdahl and Ray Jackendoff. 1983. *A Generative Theory of Tonal Music*. MIT Press, Cambridge, Mass.
[10] Fred Lerdhal. 2001. *Tonal Pitch Space*. Oxford University Press, New York.
[11] Alan Marsden. 2005. Generative structural representation of tonal music. *Journal of New Music Research* 34, 4 (2005), 409–428. https://doi.org/10.1080/09298210600578295
[12] Alan Marsden. 2010. Recognition of Variations Using Automatic Schenkerian Reduction. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*. International Society for Information Retrieval, 501–506.
[13] Yoko Oura and Giyoo Hatano. 1991. Identifying Melodies from Reduced Pitch Patterns. *Psychologica Belgica* 31, 2 (1991), 217–237.
[14] David Rizo and Alan Marsden. 2016. A Standard Format Proposal for Hierarchical Analyses and Representations. In *Proceedings of the 3rd International Workshop on Digital Libraries for Musicology (DLfM 2016)*. ACM, New York, NY, USA, 25–32. https://doi.org/10.1145/2970044.2970046
[15] Heinrich Schenker. 1979/1935. *Free Composition (Der Freie Satz)*. Longman, New York.
[16] Claude E. Shannon and Warren Weaver. 1949. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.
[17] John A. Sloboda and David H. H. Parker. 1985. Immediate Recall of Melodies. In *Music Structure and Cognition*, Ian Cross Peter Howell and Robert West (Eds.). Academic Press, London, UK, 143–167.
[18] Satoshi Tojo and Keiji Hirata. 2013. Structural Similarity Based on Time-Span Tree. In *From Sounds to Music and Emotions: 9th International Symposium CMMR, London, UK, June 19-22, 2012, Revised Selected Papers*, Richard Kronland-Martinet Mitsuko Aramaki, Mathieu Barthet and Sølvi Ystad (Eds.). Springer Verlag, Berlin, Heidelberg, 400–421.
[19] Satoshi Tojo, Alan Marsden, and Keiji Hirata. 2017. On Linear Algebraic Representation of Time-span and Prolongational Trees. In *Proceedings of the 13th International Symposium on Computer Music Multidisciplinary Research (CMMR)*. Matosinhos, Portugal.