EXTENDING A NETWORK-OF-ELABORATIONS REPRESENTATION TO POLYPHONIC MUSIC: SCHENKER AND SPECIES COUNTERPOINT

Alan Marsden
Music Department, Lancaster University
Bailrigg, Lancaster, LA1 4YW, UK
A.Marsden@lancaster.ac.uk

ABSTRACT

A system of representing melodies as a network of elaborations has been developed, and used as the basis for software which generates melodies in response to the movements of a dancer. This paper examines the issues of extending this representation system to polyphonic music, and of deriving a structural representation of this kind from a musical score. The theories of Heinrich Schenker and of Species Counterpoint are proposed as potentially fruitful bases.

1. BACKGROUND

A common theme of music theory from the eighteenth century has been that underlying the pattern of notes which forms the 'surface' of a piece of music is a less elaborate framework. The idea finds its fullest exploitation and culminating exposition in the work of Heinrich Schenker, whose seminal work *Der freie Satz* [18] had enormous influence on the music theory of the late 20th century. Computational implementations of the theory are found in the work of Kassler [9], Frankel, Rosenschein & Smoliar [8], and a number of more recent authors. Pursuing the common parallel between music and language, the theory has been compared to generative grammar, and a number of computational implementations of musical grammars have been reported also, some more closely related to Schenkerian theory (e.g., Baroni [2], and Baroni, Dalmonte & Jacoboni [3]), and others of a very different nature (e.g., Kippen & Bel [10]). The parallel with language and ideas from Schenkerian theory have come together also in the influential work of Lerdahl and Jackendoff [11], which has itself been subject to attempts at computer implementation (e.g., Baker [1]).

At the same time, it is a common finding of researchers in automatic processing of musical information (such as the 'optical music recognition', see Ng & Cooper [15]) that information about the structure of the music facilitates or is even essential for the task. Large-scale research projects have been explicitly concerned with computational accounts of musical structure (e.g. Cambouropoulos [4]), and some representations schemes have explicitly incorporated elements of structure (e.g., the Charm system of Wiggins & Smaill [20]).

Combining these various strands, a representation scheme has been developed which explicitly represents melodies not as a sequence of notes but as a network of elaborations which generate the notes of the surface of a melody from a much simpler background sequence. The scheme has been demonstrated to facilitate the representation of pattern in Music (Marsden [13]), and it has been used as the basis of software which generates melodies in response to gestural input from a dancer (Marsden [14]; the gesture-recognition component is based on EyesWeb, Camurri et al. [5]). In both cases, representation in an elaboration-network facilitates musically significant judgements and operations. In the first, patterns may be clear to the ear but not evident in the simple sequence of notes, because the surface intervals vary from one occurrence of the pattern to another or because one sequence may be a variation of another. Explicitly representing the quasi-Schenkerian structure allows the common underlying pattern to be recognised at levels below the surface. In the second, the generation of melodies by elaboration of a simple background ensures melodies that are always 'correct' (in the same way that a generative grammar is guaranteed to produce grammatical sentences), and the creation of segments of melody which are recognisably related to an earlier segment but more or less elaborate. Software which implements the representation scheme may be found at http://www.lancs.ac.uk/staff/marsdena/software/novagen.

To be properly useful, however, two important features need to be added to the representation scheme: the representation of not just melodies but polyphonic music, and the derivation of a representation from a 'surface' collection of notes. This paper discusses work in progress towards the realisation of these two missing features. They are deliberately combined in the belief that solution of one problem requires solution of the other. It is a common finding that ignoring one characteristic complicates derivation of another (e.g., determining the spelling of pitches is more difficult if rhythm is ignored than if it is taken into account). Many melodies, while presenting a single sequence of notes on the surface, actually have an underlying polyphonic structure. It would therefore be foolish to attempt automatic derivation of the structure of these melodies without first solving the problem of representing polyphonic music. A third significant issue is the basis for the definition of the vocabulary of elaborations from which a representation may be composed.

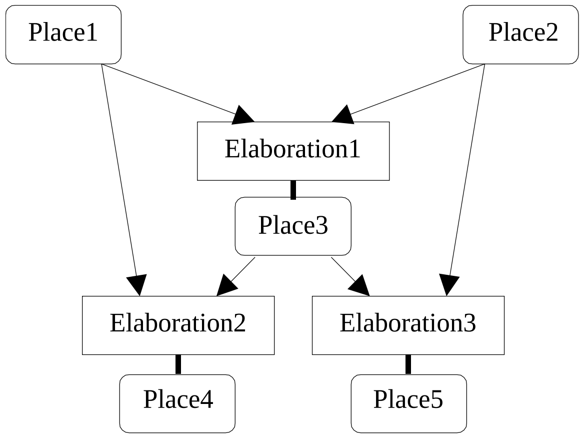

Figure 1: Basic network structure

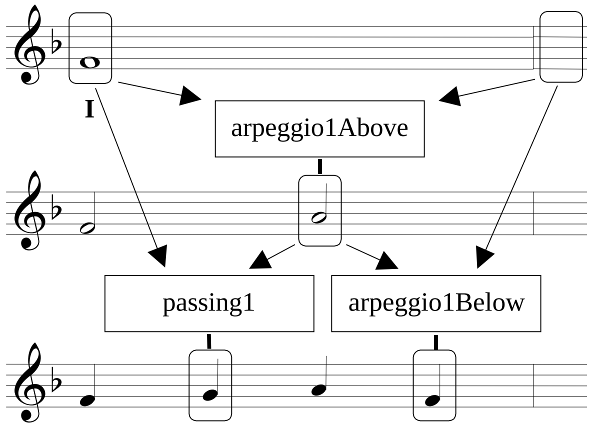

Figure 2: Representation of the beginning of Frère Jacques

2. THE FUNDAMENTAL ELABORATION-NETWORK REPRESENTATION

In the scheme as originally developed, a representation is a planar directed acyclic graph with two classes of node. 'Places' are notes or rests, and have properties of pitch (except for rests), time (but not duration) and prevailing key, harmony and metre. At the highest level, places form a sequence, with each note or rest lasting until the next begins. (Thus a dummy rest is required at the end of a melody.) 'Elaborations' are connected to two parent places, and generate a new place which occurs in time between the two parent places. The first parent now lasts only until the newly generated place, which in turn lasts until the beginning of the second parent. The pitch of the generated note depends on the kind of elaboration, the pitch of one or both of the parents, and the prevailing key and harmony. An 'upper neighbour note' elaboration, for example, produces a pitch which is one step in the scale of the key above the pitch of the second parent. The time of the generated note depends on the times of the two parents, the prevailing metre, and the nature of the elaboration, which can be 'even', 'long-short' or 'short-long'. This

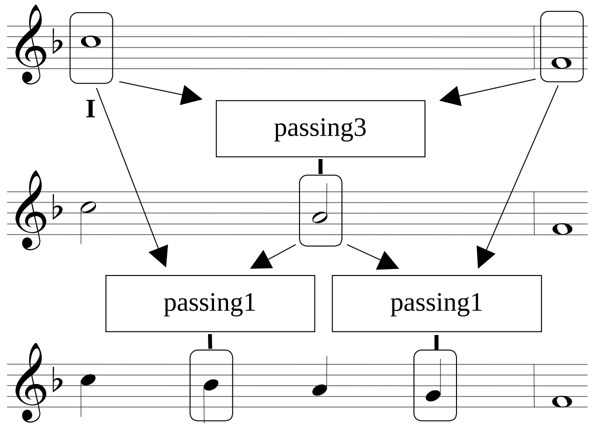

Figure 3: Passing notes spanning a fifth; original representation

new place can in turn become a parent for further elaborations. Figure 1 illustrates this basic network idea, and Figure 2 shows the representation of a fragment of a simple melody. The fundamental value of the scheme is that a melody is represented by the top-level root places and the pattern of progressive elaborations; the details of pitch and time of most of the notes of the melody are not represented directly but emerge from those elaborations. The essential premise is that, for many musical manipulations, the fundamental information required is the pattern of elaborations, and not the details of pitch and time.

Some elaboration types, however, do not fit into this simple scheme. There are two kinds of problematic case which are discussed in further detail below. In one kind, more than one new note is generated. This is most obvious in passing elaborations. While it is most common for passing notes to occur singly between two notes a third apart, pairs of passing notes between notes a fourth apart are also common. There are also cases where passing notes occur across larger intervals and while it is in principle possible to break these down into combinations of arpeggiations and passing notes (since a proper arpeggiation never produces an interval larger than a fourth), this does rather misrepresent the music in many cases because there is no sense in which the notes of the putative underlying arpeggio are distinguished from the passing notes which make up the rest of the musical gesture. In the original scheme, the simplicity of the network structure was preserved at the cost of complicating the definitions of elaborations and the dependencies between them: a group of passing notes is represented as a set of elaborations, each generating a single note. Thus a pair of passing notes spanning a fourth is represented as two elaborations: a 'passing2' elaboration which generates the passing note, and a 'passing1' elaboration between that newly generated note and the next. A set of passing notes spanning a fifth is represented by three elaborations, again each generating a single note, as shown in Figure 3.

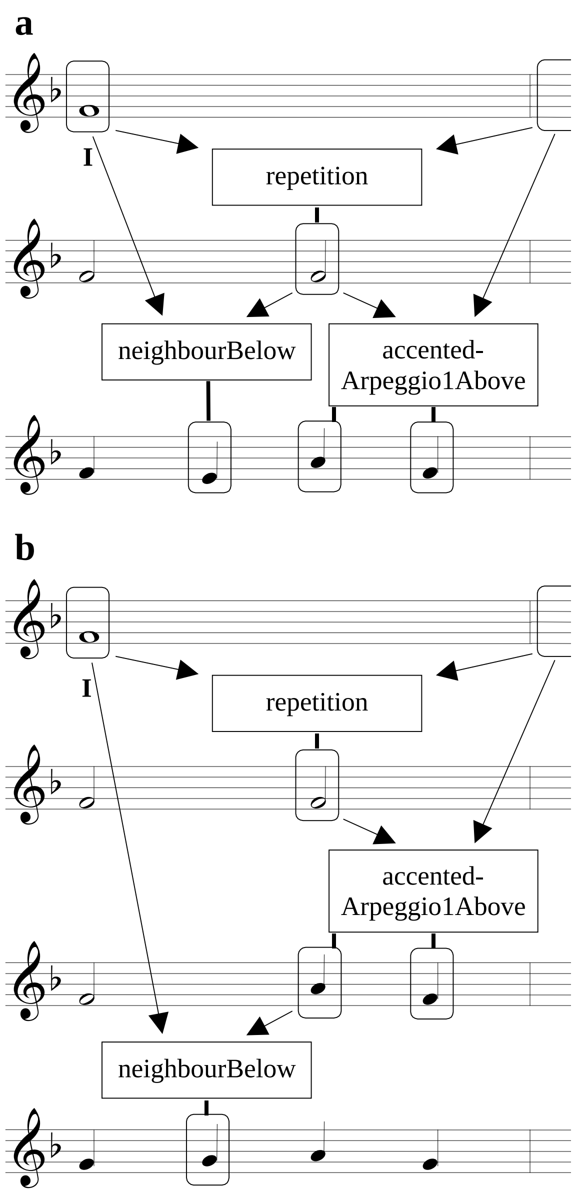

Figure 4: Two ways of attaching to an accented elaboration.

The second kind of problematic case is those elaborations where a new note replaces the elaborated note in time, while the elaborated note is generally shifted to occur later: suspensions, appoggiaturas, and the like. In the original representation scheme, these were classed as 'accented elaborations' which generated two places. The first newly generated place has the time of the original first parent while the second newly generated place has a time between then and the second parent, generally inheriting its pitch and other properties directly from the first parent. This complicates the original simple network structure firstly because an accented elaboration has two children, and secondly because places at background and middleground levels do not necessarily occur at the surface of a melody, or at least not at their original times. Most importantly, it changes the context for elaborations occurring immediately beforehand, and it becomes crucial whether these have the original first parent for their (second) parent or the newly generated note. This is illustrated in Figure 4 which shows two different representations arising from two ways of connecting the same elaboration.

The set of elaborations was determined largely intuitively on the basis of what seemed to the author necessary for the representation of common-practice tonal melodies, taking the music of the Classical era as central. The full set is given precise definition in Marsden [13], though some adaptations have been made in the on-line example software referred to above and in the melody-generation system described in Marsden [14].

3. REQUIREMENTS FOR EXTENSION

Three issues have been identified for further development of the representation system before it can be a widely useful tool for musical computing:

1. the representation of polyphonic music;
2. a process to derive a representation from a musical surface; and
3. a sound basis for the definition of the vocabulary of elaborations.

These are issues of fundamentally different kinds. Issue 1 is about the basic structure of the representation network. Issue 2 is about the development of an algorithm, but features of the representation might hinder or facilitate the development of a suitable algorithm. Issue 3 is about the epistemological basis of the representation system and its validity. Despite these fundamental differences, they might best be treated together, as argued above.

The first step should be to clarify some formal requirements, and this may best be done by considering the two 'directions' of generation of a musical surface from a representation and derivation of a representation from a surface. Some explicit consideration of epistemology is also useful.

3.1. Generation

The first and fundamental requirement of a representation is that it should generate a unique musical surface. The 'symbolic level' here roughly corresponds to the notes in a musical score. Different performances of the same piece can, of course, differ significantly, and so more than one 'surface' in sound is possible, but these differences are here regarded as 'sub-symbolic' and not to be reflected in a difference in the representation. Different printed editions of the same piece can differ in their layout of the notes, which is not significant, or in the actual sequences of notes, which is significant. These latter differences should be reflected in different representations. On the other hand, the same piece might have more than one representation (as is clear from

Figures 2 and 4), corresponding to different interpretations of its structure. In summary, the relation between a representation and a piece of music should be many-to-one. The 'surface' of a representation in the original definition is that sequence of places which consists of all the places of the representation which have not been replaced by accented elaborations. This process of deriving this surface sequence will be called 'realisation', and in the original version it is clearly unique. It is important that extensions do not destroy this property.

A second important requirement is that realisation should be efficient, which in practice means that there must exist an algorithm to generate the surface whose complexity is of less than polynomial order in relation to the size of a representation, and preferably of not significantly greater than linear order. For example, if the representation is to be used in a live performance system, a realisation algorithm of linear complexity means that if a piece can be 'performed' in real time, then an extension of that piece can most likely also be performed in real time: the increase in the size of the representation will be accompanied by a corresponding increase in the time available to perform it. If the algorithm is of a greater order of complexity, then continually extending the piece will inevitably lead to a point where the piece can no longer be performed in real time.

The original representation scheme fulfils this second requirement because the information required for realisation was always locally available: realising one segment of the network does not require reference to a distant portion, and so realising that segment within a large network does not take any longer than realising it within a small network, and the time required to realise an entire representation grows linearly with the size of the network. The fact that the representation is a network rather than a tree means that it is not simply decomposable: extracting a subnetwork might lose some information required for proper realisation through the links which must be severed. However, because the network is planar, the information lost is confined to the edge of the region from which the subnetwork is extracted, so the representation is just one step away from being decomposable, and little needs to be added to an extracted portion to replace the lost information. It is important that this property should be retained in extension of the representation scheme.

3.2. Derivation

As pointed out above, the derivation of a representation from a musical surface need not necessarily be unique. However, some representations are almost certainly better than others, and so a derivation algorithm should ideally find a representation which is optimal in some sense. What 'optimal' might mean here is considered briefly below, but must be a major topic for research.

Once again, to be of use, derivation must be possible with reasonable efficiency. An algorithm of polynomial order might be satisfactory for analytical tasks, but if the representation is to be used in real-time systems, once again the algorithm must approach complexity of linear order. If, for example, an automatic ensemble improvisation system is to use the representation to act on the basis of the structure of the music played by the other members of the ensemble, it must be able to derive the structure about as quickly as the music is played, regardless of its length. Similar considerations require the algorithm to be bottom-up: a top-down derivation algorithm would require the whole piece to be complete before any derivation of structure could begin.

As with realisation, the key to efficient derivation is decomposability: to what degree is it possible to derive the optimal representation of a segment of the piece without reference to the remainder of the piece which forms the context of that segment? The answer to this is that it must be possible to find a representation for segments of musical surface which correspond to those segments of a piece which are decomposable in realisation. However, while that representation might be optimal for the segment extracted from the context of the entire piece, it might not be optimal in its proper context. Complete decomposability is unlikely to be possible, and some reference to music outside the local context is almost inevitable. Two particular kinds of influence from the non-local context are to be expected. Assuming that an aspect of what makes a representation 'optimal' is that it maximises the use of recurrent patterns, determination of the optimal representation of a segment is likely to require reference to the representations of earlier segments in order to recognise whether the current segment repeats a pattern from an earlier segment or not. Since in a real-time application, the representation of these segments will always already have been derived, this does not contradict the benefit of a bottom-up derivation algorithm noted above. However, if the derivation algorithm requires comparison of the current segment with every previous segments, the number of comparisons required for later segments would increase with the increasing length of the piece to be represented, and an algorithm of quadratic order would be required. Thus a more subtle pattern-tracking approach is to be preferred. An additional complication arises from the possibility that maximising the use of recurrent patterns might require the representation of an earlier segment to be revised in the light of a later segment which is similar but not identical. (A rather complex case of this kind of behaviour in Mozart's string quartet in C major, K465, the 'Dissonance', is discussed in Marsden [12], where it is suggested that this phenomenon is one of the features of the music which retains the listener's interest in repeated hearings.)

The other kind of non-local reference expected to be required for deriving an optimal representation arises from the possibility that optimal representation of higher levels in a structure might require representations at lower levels which would be less than optimal if those lower-level segments were taken out of context. In other words, in a bottom-up approach, decisions might be taken early on which appeared optimal at that stage but which prevent optimal representations later on at a higher level.

The kinds of decisions required at each stage are basically of two kinds: which notes of the surface are to be regarded as of the lowest level (i.e., the ones generated by the lowest-level elaborations); and what kind of elaboration is to be regarded as generating those notes. In the smallest possible local segment, the number of possibilities of each kind will be small (in single figures), but the number of possible combinations of course increases exponentially as the size of the segment increases. The point made above that optimal representation at higher levels cannot be guaranteed from optimal representation at lower levels means that an algorithm which simple pursued the locally optimal decision at each point would be unlikely to produce an optimal representation for the entire piece. That the number of possible representations increases exponentially with the size of the piece means that an algorithm which pursues every possible decision at each point is unworkable. Thus a good pruning regime is essential, which at each stage identifies and cuts off branches of possibilities which cannot lead to an ultimately optimal representation but keeps alive possibilities which might later lead to a better representation than the one derived so far.

3.3. Epistemology

It has often been argued that, especially in the domain of music, computational processes are most appropriate when they mimic in some way human processing. What might the psychological validity of the representation scheme be? What principles from music psychology might guide its extension? While often examining the question, psychologists have rarely been so bold as to propose how music is actually represented in the mind. One exception is the scheme of Deutsch and Feroe [7], which shares a great deal with this representation scheme: it is hierarchical—simpler background sequences are elaborated into more complex sequences—and it makes use of the ‘alphabets’ of scales and arpeggios. Other work has shown in a number of ways that listeners extract the structural features of the music they hear, being particularly sensitive to pattern (e.g., Deutsch [6] and Sloboda & Parker [19]). However, while giving support to the general idea, none of this work has been sufficiently fine-grained to provide guidance for the detail of a representation scheme for polyphonic music.

It was mentioned above that the original representation scheme was inspired by Schenkerian theory. While not having the scientific force of psychological research, this does provide a degree of support to the representation scheme. In its adoption by so many music theorists and analysts, Schenkerian theory has proven itself to be useful for the discussion of musical structure at the level of abstract analysis. It can therefore be expected to provide a source of guidance for the development of a scheme for the representation of polyphonic music. In particular, Schenker identifies a number of manners of ‘diminution’ which could be translated into kinds of elaboration in the representation scheme. Although Schenker does not discuss the process of analysis much himself, later research, particularly by Schachter [17] and Plum [16], has examined the principles on which decisions are made in deriving an analysis from a score, and these (especially Plum’s ‘indices’) might usefully guide the development of that part of a derivation algorithm which selects optimal representations. Finally, and perhaps most importantly, if Schenkerian graphs can be directly related to elaboration-network representations, then Schenker’s own analyses and those of his many followers provide a rich test corpus.

Species counterpoint holds validity by the same kind of utilitarian argument as applies for Schenkerian theory—indeed more so because of its longer history, and Schenker claimed his own theory to be based on species counterpoint. This theory does not represent pieces hierarchically and it does not encompass an explicit account of elaboration (unlike theories of musical ornamentation, the other historical source of Schenker’s theory), but its rules of counterpoint can be related to the constraints which apply to elaborations and, most usefully, it does give an account of counterpoint broken down into its simplest components. In particular, it identifies a small number of fundamental classes of contrapuntal configuration, from which all proper counterpoint is constructed, and it reduces counterpoint in any number of voices to the relationships between pairs of voices. These are likely to provide useful guidance for the development of a representation scheme for polyphonic music.

4. PROPOSALS

4.1. Dealing with simultaneous voices

The original representation scheme could represent some polyphonic music simply by allowing multiple places to coexist at the same time. The splitting of one voice into two and the joining of two voices into one could be accommodated by allowing places to be simultaneously elaborated by more than one elaboration (one generating one voice and one another). There are, of course, harmonic constraints on such simultaneous notes and elaborations, and it is important that these do not lead to interdependencies between elaborations which would render either the realisation process or the derivation process intractable. However, the constraints generally concern validity rather than the detail of interpretation. For example, in most tonal music, simultaneous passing notes are only allowed when they are consonant with each other. In realisation, the presence of other passing notes should not cause a passing-note elaboration to generate different notes than it would if those other passing notes were not present. In derivation, the presence of other passing notes will be one factor in determining whether or not a passing-note elaboration is a valid way of representing the music at this point. (One counter example might be the practice of ‘musica ficta’ in Mediaeval and Renaissance music, when a note might be sharpened or flattened to avoid a tritone interval with a note in another part. To accommodate such instances under these proposals would require different types of elaborations for the ‘ficta’ and ‘recta’ versions (roughly ‘chromatic’ and ‘diatonic’ according to modern terminology), but this might imply a difference in melodic pattern which would not have been considered significant at the time.)

This proposal therefore does not complicate realisation, though it does complicate derivation. On the other hand, derivation is likely to be complicated anyway. One alternative which has been considered is to allow places to contain more than one simultaneous note, and to have polyphonic elaborations act on such ‘chordal’ places to generate simultaneous voices. However, this alternative has two apparent disadvantages. The first is that the number of elaboration types required is likely to be very large in order to accommodate all the kinds of polyphonic configurations which are encountered in pieces of music. The second is that this does not allow easy decomposition of a polyphonic representation into monophonic representations of separate voices. The possibility of decomposition was seen earlier to be an important property leading to tractability, and so should be retained if possible.

Some additional elaborations types will required to accommodate some specific polyphonic configurations, in particular the pattern Schenker calls ‘unfolding’, where notes, which at a deeper level make up a multi-voice sequence moving in diads or chords, are presented at the surface alternating in a single voice (resulting in the phenomenon sometimes called ‘pseudo-polyphony’ or a ‘compound voice’). Such elaborations will need to have as parents two (or more) simultaneous notes, which further complicates the network structure of a representation.

4.2. An alternative approach to elaborations

On a number of occasions above, complications introduced by the possibility of accented elaborations, i.e., those which replace the first parent by some other note, have been mentioned, and of multiple-note elaborations like passing notes over an interval greater than a third. To avoid these complications, is proposed to change the definition of an elaboration in two ways (which both have the effect of simplifying the possible structures in a representation in some ways but complicating it in others!).

Firstly, it is proposed that elaborations should have a single parent rather than two as in the original scheme. The single parent will correspond to the left parent of the original scheme, and all new notes generated by the elaboration will occur in the time span of that parent note (and will fill the time span). For some elaborations, this is all the information required. (A repetition, for example, requires no information other than the time span and pitch of the parent note.) In the simplest cases, therefore, a representation will be a tree structure similar to the ‘time span reductions’ of Lerdahl & Jackendoff. (The one exception to this will be polyphonic elaborations, such as the unfoldings referred to above, which will have multiple parents, but these will all share a single time-span.)

Some elaborations, on the other hand, require information beyond that of the time-span in which the new notes should be generated and the pitch of the note currently occupying that span. Realisation of a passing-note elaboration, for example, requires knowledge of the pitch of a following note. Realisation of a suspension elaboration requires knowledge of a preceding note. I say ‘a … note’ rather than ‘the … note’ here because while it is evident that the following or preceding note must occupy a time-span adjacent to the one elaborated, in polyphonic music there might be several such notes in valid positions. It is proposed, therefore, that an elaboration should, where necessary, have links to the appropriate following or preceding note to provide the proper contextual information for generation of the new note(s). These links will link adjacent branches of the simple tree structure referred to above, so resulting in a network with the properties similar to the original network scheme described above. Because links will always be to an immediately preceding or following time-span, links will never actually ‘cross’ a branch, but it is not impossible in a polyphonic piece that links might cross each other. This will have consequences for the decomposability of a representation.

Finally, it is proposed that elaborations should replace their parent (or parents in the case of polyphonic elaborations) with a sequence of notes, which in most cases will be just two notes long but, in cases such as passing notes and arpeggios, could be longer (indeed, of unbounded length, in principle). The sequence produced must contain a note (or notes) with the same pitch (and other properties, if appropriate) as the parent (or parents). This rules out the possibility of substitution elaborations, such as Schenker considers possible in certain cases [18, p.51], but these are rare, and allowing them enormously complicates derivation. (In the worst case, it would lead to the possiblity of circular recursion and hence infinite derivation.)

5. REFERENCES

[1] Michael Baker. A Computational Approach to Modeling Musical Grouping. *Contemporary Music Review*, 4, 1989, 311–325.

[2] Mario Baroni. The Concept of Musical Grammar. *Music Analysis*, 2, 1983, 175–208.

[3] Mario Baroni, Rossana Dalmonte & Carlo Jacoboni. Theory and Analysis of European Melody. In Alan Marsden & Anthony Pople (eds.) *Computer Representations and Models in Music*, London: Academic Press, 1992, 187–205.

[4] Emilios Cambouropoulos. Towards a General Computational Theory of Musical Structure, PhD thesis, University of Edinburgh, 1998.

[5] A. Camurri, S. Hashimoto, M. Ricchetti, R. Trocca, K. Suzuki, G. Volpe. EyesWeb – Toward Gesture and Affect Recognition in Interactive Dance and Music Systems. *Computer Music Journal*, 24, 2000, 57–69.

[6] D. Deutsch. The Processing of Pitch Combinations. In D. Deutsch (ed.) *The Psychology of Music* (2nd edition), San Diego: Academic Press, 1999, 349–411.

[7] D. Deutsch & J. Feroe. The Internal Representation of Pitch Sequences in Tonal Music. *Psychological Review*, 88, 1981, 209–220.

[8] R.E. Frankel, S.J. Rosenschein & S.W. Smoliar. A LISP-Based System for the Study of Schenkerian Analysis. *Computers and the Humanities*, 10, 1976, 21–32.

[9] Michael Kassler. A Trinity of Essays. PhD thesis, Princeton University, 1967.

[10] Jim Kippen & Bernard Bel. Modelling Music with Grammars: Formal Language Representation in the Bol Processor. In Alan Marsden & Anthony Pople (eds.) *Computer Representations and Models in Music*, London: Academic Press, 1992, 207-238.

[11] Fred Lerdahl & Ray Jackendoff. *A Generative Theory of Tonal Music*. Cambridge, Mass.: MIT Press, 1983

[12] Alan Marsden. Listening as Discovery Learning. *Contemporary Music Review*, 4, 1989, 327–340.

[13] Alan Marsden. Representing Melodic Patterns as Networks of Elaborations. *Computers and the Humanities*, 35, 2001, 37-54.

[14] Alan Marsden. Novagen: A Combination of Eyesweb and an Elaboration-Network Representation for the Generation of Melodies under Gestural Control. In *Proceedings of the COST287-ConGAS Symposium on Gesture Interfaces for Multimedia Systems (GIMS)*, Leeds University, 2004.

[15] Kia C. Ng & David Cooper. Enhancement of Optical Music Recognition using Metric Analysis, in *Proceedings of the XIII CIM 2000—Colloquium on Musical Informatics*, L'Aquila, Italy, Sep 2000, 189–192.

[16] Karl-Otto Plum. Towards a Methodology for Schenkerian Analysis (trans. and ed. William Drabkin), *Music Analysis*, 7, 1988, 143–164.

[17] Carl Schachter. Either/Or. In C. Schachter (ed.) *Unfoldings*, Oxford: Oxford University Press, 1999, 121–133.

[18] Heinrich Schenker. *Der Freie Satz*. Vienna: Universal Edition, 1935. Published in English as *Free Composition*, translated and edited by Ernst Oster. New York: Longman, 1979.

[19] John A. Sloboda & David H.H. Parker. Immediate Recall of Melodies. In P. Howell, I. Cross & R. West (eds.) *Musical Structure and Cognition*, London: Academic Press, 1985, 143–167.

[20] Geraint Wiggins & Alan Smaill. Musical Knowledge: What can Artificial Intelligence Bring to the Musician? In Eduardo Reck Miranda (ed.) *Readings in Music and Artificial Intelligence*, Amsterdam: Harwood Academic Publishers, 2000, 29–46.