PROBABILISTIC MODELING OF HIERARCHICAL MUSIC ANALYSIS

Phillip B. Kirlin and **David D. Jensen**
Department of Computer Science, University of Massachusetts Amherst
{pkirlin,jensen}@cs.umass.edu

ABSTRACT

Hierarchical music analysis, as exemplified by Schenkerian analysis, describes the structure of a musical composition by a hierarchy among its notes. Each analysis defines a set of prolongations, where musical objects persist in time even though others are present. We present a formal model for representing hierarchical music analysis, probabilistic interpretations of that model, and an efficient algorithm for computing the most probable analysis under these interpretations. We represent Schenkerian analyses as maximal outerplanar graphs (MOPs). We use this representation to encode the largest known data set of computer-processable Schenkerian analyses, and we use these data to identify statistical regularities in the human-generated analyses. We show that a dynamic programming algorithm can be applied to these regularities to identify the maximum likelihood analysis for a given piece of music.

1. INTRODUCTION

Schenkerian analysis [13] is a widely used and well-developed approach to music analysis. Analyses interpret compositions as a hierarchical structure of musical events, allowing a user to view a tonal composition as a collection of recursive musical elaborations of some fundamental structure. The method of analysis starts from the original composition and produces a sequence of intermediate analyses illustrating successive simplifications or *reductions* of the musical structure of the piece, ultimately arriving at an irreducible background structure. Each reduction is a claim that a group of musical events (such as notes, intervals, or harmonies) X derives its function within the composition from the presence of another group of events Y, and therefore the overarching musical structure of the collection $X \cup Y$ is determined predominantly by the events in Y. In Schenkerian terms, we often say the events in X constitute a *pro-*

longation of the events in Y, in that the events in Y remain "in effect without being literally represented at every moment." [2]

Schenker's ideas may be viewed as a set of tools for constructing a hierarchical analysis of a composition according to the analyst's own musical intuition, or as theory of tonality such that every tonal composition, and only tonal compositions, should be derivable from the "rules" of Schenkerian analysis [1, 15].

Opinions differ about the underlying goals of Schenkerian analysis. However, one thing is clear: Schenker's ideas alone do not prescribe an unambiguous and complete algorithm for analysis. That said, generations of music theorists have used Schenker's ideas to construct analyses. In this paper, we pursue an empirical strategy for discovering the underlying regularities of those analyses and producing new analyses based on those regularities. Specifically, we derive statistical regularities from the largest known corpus of machine-readable Schenkerian analyses, and we identify an algorithm for deriving the maximum likelihood analysis, given these regularities. We demonstrate that the algorithm can reproduce the likelihood ranking implied by a probability distribution over possible analyses. Together, these findings provide the foundation of an empirical strategy for unlocking the basic concepts underlying any method of hierarchical music analysis.

2. REPRESENTATIONS AND ALGORITHMS FOR SCHENKERIAN ANALYSES

Tree-like data structures are natural representations for hierarchies. Combined with the Schenkerian idea of the analysis procedure revealing multiple levels of musical structure in a composition, many researchers have used different types of trees to represent an analysis and the structural levels within.

A commonly used tree representation of an analysis uses leaf nodes to represent notes or chords of the original composition, and interior nodes to represent Schenkerian reductions of each node's children. This formulation has been used by Frankel, Rosenschein and Smoliar [3,4], Rahn [12], Lerdahl and Jackendoff [8], Marsden [9, 10] and Kirlin [7]. Algorithms for analysis that use such representations have had varying levels of success [6, 10, 11, 14].

Yust argues for using a hierarchy of melodic intervals—the spaces between the notes—rather than the notes or chords themselves. He contends that such a hierarchy of intervals better reflects Schenker's original ideas and reduces the size of the search space of analyses [15]. Mavromatis and Brown [11] and Gilbert and Conklin [5] also suggest an interval-based hierarchy would alleviate some representational problems.

Consider Figure 1(a), an arpeggiation of a G major triad with passing tones between the notes of the chord. Representing this musical figure as a hierarchy of notes forces us to choose a single parent note for each passing tone, obscuring the nature of a passing tone as a voice-leading connection from one note to another. Using a hierarchy among intervals between the notes, however, allows us to represent the musical structure as the tree in Figure 1(b). If we then replace the nodes of this tree with edges, we obtain the representation in Figure 1(c), a particular kind of graph called a *maximal outerplanar graph*, or MOP, a representation for musical analysis first suggested by Yust [15]. MOPs are isomorphic to binary trees representing interval hierarchies such as that in Figure 1(b), though because the MOP does not duplicate notes as the tree does, it is a more compact representation of the hierarchy.
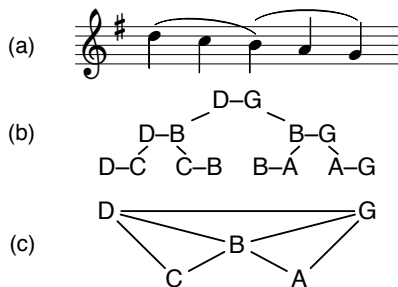

Figure 1. (a) An arpeggiation of a chord with passing tones. (b) A hierarchy among the melodic intervals in the arpeggiation. (c) The MOP corresponding to the arpeggiation.

Every MOP defined on a given sequence of notes is a triangulation of the polygon formed by the edges between consecutive notes and the edge from the first note to the last note. Each triangle in the MOP specifies a prolongation among three notes; we will occasionally refer to a triangle as containing two *parent notes* and a single *child note*, or a single *parent interval* and two *child intervals*. Either interpretation is musically correct: the left parent note is prolonged by the child note during the time span between the left and right parent notes, or the melodic interval between the left and right parent notes is prolonged by the motion to and away from the child note.

Because every prolongation requires two parent notes, incomplete prolongations, such as incomplete neighbor notes, present a representational challenge in MOPs. Yust argues

that in these situations, it is appropriate to have the nearest structural note substitute for the missing parent note. To allow for incomplete prolongations at the beginning or ending of a piece, the MOP model places special "initiation" and "termination" events at the beginning and ending of the passage being analyzed that may be used as parents for such prolongations.

The MOP model offers a new look at representation of analyses that more closely parallels Schenkerian analysis in practice due to the MOP's emphasis on preserving voice leading connections. Further discussion of MOPs may be found in Yust's dissertation [15].

3. A GENERALIZATION OF MOP: OPC

The definition of a MOP stated above can only handle a single monophonic sequence of notes, though the model can be extended to allow for a single structure to represent the analysis of a contrapuntal or polyphonic composition [15]. However, in the interest of simplicity, we have chosen to store such analyses as collections of separate MOPs occurring simultaneously in time. For instance, in a two-voice composition, there would be one MOP to represent the upper voice, and one MOP to represent the lower voice. Taking both MOPs together as a collective representation of an analysis gives us an *OPC* (outerplanar graph collection).

The OPC representation also relaxes one restriction on the constituent MOPs, namely that the polygon formed by the edges connecting the notes of the composition must be completely triangulated. This is allowed because many analyses done by humans contain prolongations with multiple child notes. Such prolongations must necessarily be represented by polygons larger than triangles; in general, a prolongation with n children will be represented in an OPC by a polygon with $n + 2$ sides.

We devised a text-based file format that can encode many of the annotations found in a Schenkerian analysis, including any type of prolongation (such as passing tones, neighbor tones, and similar diminutions), voice exchanges, verticalizations of notes, repeated notes merged in an analysis, and instantiations of the *Ursatz* (the fundamental background structure posited by Schenker). The format is easy for the human to input and easy for the computer to parse. We also developed an algorithm to convert an analysis in this encoding into an OPC.

4. EXPLORATION OF ANALYSES AS MOPS

We collected a set of eight excerpts of music along with Schenkerian analyses of the excerpts. The excerpts and analyses were drawn from Forte and Gilbert's *Introduction to Schenkerian Analysis* [2] and the accompanying instructor's manual, and were chosen for their similar characteristics: they are all from compositions for a keyboard instrument in

a major key, do not modulate within the excerpt, and have a complete instance of the *Ursatz*, possibly with an interruption. The analyses were algorithmically translated to OPCs. The data set contained 66 measures of music and 617 notes. Overall, 270 prolongations were translated into 356 polygons in the OPCs. Though small, this corpus represents the largest known data set of machine-readable Schenkerian analyses. [1]

Because we are interested in prolongational patterns and each triangle in a MOP specifies the prolongation of an interval by two other intervals, we examined how often certain types of triangles occurred in the human-produced analyses represented as OPCs. We defined a triangle by an ordered triple of the size of the parent interval and the sizes of the two child intervals. Intervals were denoted by size only, not quality or direction (e.g., an ascending major third was considered equivalent to a descending minor third), except in the case of unisons, where we distinguished between perfect and non-perfect unisons. Intervening octaves in intervals were removed (e.g., octaves were reduced to unisons), and furthermore, if any interval was larger than a fourth, it was inverted in the triple. These transformations equate prolongations that are identical under octave displacement.

Because OPC analyses permit polygons larger than triangles, extra care was required to derive appropriate triangle frequencies for these larger polygons. As any polygon can only be triangulated in a fixed number of ways, and each of those triangulations contains the same number of triangles, for every polygon larger than a triangle we counted the frequencies of every possible triangle over all possible triangulations of the polygon and weighted the resulting frequencies so that they would sum to the number of triangles expected in a triangulation.

We tested the triangle frequencies to see if they were statistically significant given the null hypothesis that the Forte and Gilbert analyses resemble random analyses (where any triangulation of a MOP is as likely as any other) in their triangle frequencies. The expected frequencies under the null hypothesis are not uniformly distributed, even if all the notes in a composition are considered distinguishable from each other. Therefore, for each excerpt in our corpus, we generated 5,000 analyses of the excerpt uniformly at random. Each of these analyses was produced by taking the corresponding human-created analysis as an OPC and re-triangulating each MOP inside. We used these random analyses to compute the expected frequencies of every type of triangle possible and compared them to the observed frequencies from the human-produced analyses. We ran individual binomial tests for each type of triangle to determine if the observed frequency differed significantly from the expected frequency.

Five types of triangles had differences between their ob-

served and expected frequencies that were statistically significant at the 5% level; these are shown in Figure 2. A canonical prolongation for each type of triangle is depicted at the far left of each row in the figure, though because intervals have had intervening octaves removed and are inverted if larger than a fourth, each type of triangle represents an entire class of prolongations. Triangles that contained a perfect unison as a child interval are not shown in this table, as we suspect their frequencies are biased due to the way merged notes are encoded in an analysis. Consecutive notes of the same pitch are often implicitly merged in a Schenkerian analysis, and these are encoded as prolongations of the interval from the first note with the repeated pitch to the note following the last note with the repeated pitch.

We can musically interpret each of the five types of triangles shown in Figure 2 and hypothesize the reasons for the differences in frequency. The first row in the figure ($p = 0.001$) tells us that triangles describing an interval of a third being elaborated by two seconds are more likely to appear in a human-produced analysis than in a randomly-generated analysis. A passing tone filling in the interval of a third would fall into this category. We suspect such patterns are numerous due to the theorist's preference for identifying stepwise voice leading connections in an analysis. The second row ($p = 0.003$) shows us the commonality of a melodic second being elaborated by a third and then a step in the opposite direction, for instance, when the interval C–D is elaborated as C–E–D. Again, this corresponds to the frequent situation of a stepwise pattern being decorated by an intermediate leap. The third row ($p = 0.02$) shows the preponderance of melodic fifths (inverted fourths) being elaborated by consecutive thirds, corresponding to the arpeggiation of a triad. Harmonies are frequently prolonged by arpeggiations of this type.

The fourth row in Figure 2 ($p = 0.03$) shows that triangles corresponding to a melodic second elaborated by a step and then a leap of a third in the opposite direction occur less frequently than expected. An example would be the interval C–D being elaborated by the pattern C–B–D. Interestingly, this is the reverse case of the second row in the table. We hypothesize that analysts tend not to locate this type of prolongation because the leap of a third could suggest a change of harmony, and therefore it is more likely that the first note of the new harmony—the B in the example—would be the more structural note and not the D as would be implied by such a prolongation. The last row ($p = 0.05$) illustrates another type of prolongation found less often than the random analyses would suggest: a melodic fourth being elaborated by a step and a leap in the same direction. Musically, this type of prolongation could be located infrequently in an analysis for the same reasons as the prolongation described in the fourth row.

These statistically significant differences show that there

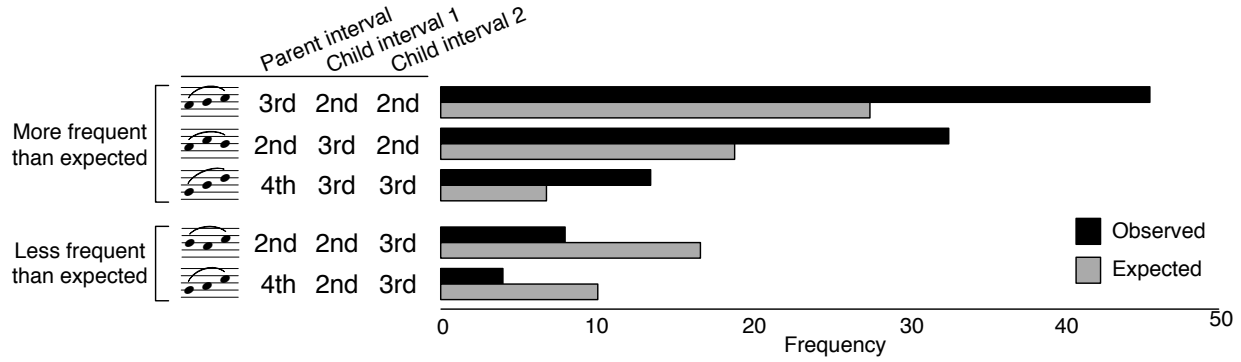[1] Analyses are available at http://www.cs.umass.edu/~pkirlin/schenker.

Figure 2. Observed and expected frequencies of triangles in the corpus of OPC analyses.

are consistencies in the prolongations that analysts locate during Schenkerian analysis. Whether those consistencies are due to the analysis method or the analyst's own proclivities is irrelevant, as the consistencies can be exploited to produce an analysis algorithm in either case.

5. PROBABILISTIC INTERPRETATIONS OF MOPS

We now show how to harness the frequencies computed in the previous section to produce an algorithm capable of hierarchical music analysis. Though we previously defined a triangle by the intervals between the notes of its vertices, in this section we will explore triangles defined by the notes themselves. Defining a triangle in this fashion requires more data than we currently have to obtain statistical significance, but we believe using this formulation will lead to better performance in the future.

With a set of triangle frequencies defined by the endpoints of the triangles in a MOP, we may define a number of different probability distributions using these frequencies. If we call the left parent note L, the right parent note R, and the child note C, we define the *joint triangle distribution* as $P(L, R, C)$. This distribution tells us the overall probability of seeing a certain type of triangle in any analysis. We also define the *conditional triangle distribution* as $P(C \mid L, R)$, which tells us the probability that the interval between the left parent note and the right parent note will be elaborated by the child note C.

Using either of these two distributions, we can define the probability of a Schenkerian analysis in the MOP model. Given that a MOP is completely defined by its constituent triangles, we define the probability of a MOP analysis for a given sequence of notes as the joint probability of all the triangles that comprise the MOP. If a MOP analysis A for a given sequence of notes N contains triangles $T_1, \ldots, T_n$, then we state $P(A \mid N) = P(T_1, \ldots, T_n)$. However, training such a joint model directly would require orders of magnitude more data than we suspect could ever be collected. Instead, as an approximation, we will assume that the presence of a certain triangle in an analysis is independent of the presence of all the other triangles. Thus, $P(A \mid N) = P(T_1) \cdots P(T_n)$.

The question remains whether to use the joint or conditional triangle distributions to define $P(T_i)$. The joint model better reflects overall frequencies of triangles, but the conditional model easily provides a generative strawman algorithm for producing an analysis: to analyze a sequence of notes $n_1, \ldots, n_k$, find $\arg\max_{i \in \{n_2, \ldots, n_{k-1}\}} P(C = i \mid L = n_1, R = n_k)$ to find an appropriate child note of n_1 and n_k, then recursively perform the same operation on the two resulting child intervals.

The issue of triangle independence remains, regardless of the specific triangle model chosen. An experiment justifies our independence assumption. Our goal in the experiment is to use a random procedure to generate a multiset of analyses for a single piece of music, with the frequencies in the multiset reflecting the real-world distribution of how analysts would interpret a piece. The ranking of the analyses by frequency in the multiset serves as ground-truth. Using this corpus of generated analyses, we compute triangle frequencies from the corpus as described in Section 4 (though using triangle endpoints instead of intervals between endpoints) and obtain a probability estimate for each analysis by using the independence of triangles assumption. We compare the ground-truth ranking with a new ranking obtained by sorting the analyses by the newly-obtained probability estimates.

The exact procedure is as follows. We assumed that every note in the piece was distinguishable from every other note, something not feasible for earlier experiments but done here with the knowledge that humans may use a note's location within the piece as a feature of the note to guide the analysis procedure. Therefore, each piece was a sequence of integers $N = 1, 2, \ldots, n$. We took a uniform sample of 1,000 MOPs from the space of possible MOPs over N. [2] We randomly chose one MOP to be the "best" analysis, and created an array A with the 1,000 MOPs sorted in decreasing order of similarity to the best MOP, where similarity was defined as the number of triangles in common between two MOPs.

---

[2] The number of MOPs for a sequence of length n is the $(n + 2)$th Catalan number [15], which is exponential in n, hence the sampling.

The best MOP was placed at $A[0]$. We used a variation of the normal distribution to sample one million MOPs from A as follows: each sample was the MOP at position i in the array, where i was the absolute value of a normal random variable with $\mu = 0$ and varying σ, rounded down. Values of i that corresponded to MOPs outside of array A were re-sampled. The one million sampled MOPs were placed into a multiset M and sorted by decreasing frequency into an array R, representing the ground-truth ranking of MOPs.

We then computed the frequency of each triangle in multiset M, calculated the probabilities for each triangle under the joint and conditional models, and used the independence of triangles assumption to compute a probability estimate for each MOP. We generated a new ranking R' of the MOPs from their probability estimates, and computed Spearman's ρ and Kendall's τ ranking correlation coefficients for R versus R' using lengths of note sequences between 10 and 50, and standard deviations σ for the normal distribution varying between 1 and 20. σ determines the number of analyses r ranked in R and R' by the formula $r \approx 4.66\sigma + 1.65$. In other words, when $\sigma = 1$, the random procedure only selects five or six analyses from the 1,000 available in A, but when $\sigma = 20$, approximately 95 are selected.

Figure 3 shows heatmaps for ρ; darker values are closer to 1, indicating R' being closer to R. The heatmaps for τ are similar. For the joint model, mean values of (ρ, τ) are $(0.9630, 0.8848)$ while for the conditional model they are $(0.9478, 0.8286)$, indicating that the joint model slightly outperforms the conditional model.
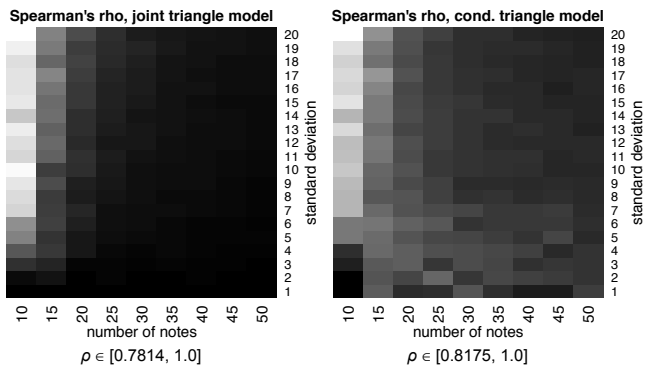

Figure 3. The joint model reproduces the ground-truth ranking slightly better than the conditional model.

Assuming independence among the triangles in a MOP provides us with an algorithm for calculating the most probable MOP, regardless of whether we choose the joint or conditional models for the probability of an individual triangle, or some other model of triangle probability. Because constructing a MOP is equivalent to triangulating a simple convex polygon, we may take advantage of the fact that this optimal triangulation problem can be solved in $O(n^3)$ time using a Viterbi-like dynamic programming algorithm where n is the number of notes in the composition. We will refer to this algorithm as OPT-MOP.

6. EVALUATION

To evaluate OPT-MOP and the suitability of the joint and conditional triangle models, we performed a leave-one-out cross-validation test. We generated 1,000 optimal analyses of the MOPs contained in each of the eight excerpts in our corpus by using, for each excerpt, triangle probabilities derived only from the ground-truth analyses of the other seven excerpts. We needed to compute multiple optimal analyses as occasionally ties appeared among the probabilities; OPT-MOP broke these ties randomly. Additionally, we generated 1,000 analyses uniformly at random for each excerpt.

To measure the quality of a candidate analysis A, we calculated the number of triangles in A that were compatible with the corresponding ground-truth analysis. We say a triangle is compatible with the ground-truth if it is present in the ground-truth (the three specific notes of the excerpt are triangulated the same way in both analyses), or if there is nothing in the ground-truth analysis that would prevent such a triangle from appearing in the ground-truth. The second provision is required because the ground-truth is human-produced and may contain prolongations that do not specify a complete triangulation. Therefore, any triangle that could result from further triangulation is deemed compatible.

We compared the mean percentage of compatible triangles in the optimal analyses with the corresponding percentage for the random analyses. Comparisons were done separately for the joint and conditional models. Table 4 shows the mean compatibility percentages under both models, along with a p-value calculated under the null hypothesis that the OPT-MOP does not perform better than random. These data indicate that both models perform better than random as a whole, because if the null hypothesis were true, we would expect only one of the eight pieces to have a p-value less than 0.1 for either model. Furthermore, the joint model outperforms the conditional model on average.

There are a number of possible reasons why the results are not better. First, the ground-truth analyses are not completely triangulated, and this puts an upper bound on how well OPT-MOP can improve over random analyses. As an extreme example, if a MOP were not triangulated at all, then all triangles produced by any analysis algorithm would be compatible with the ground-truth, and therefore both OPT-MOP's analyses and the random analyses would both obtain scores of 100%.

Second, it is not surprising that a training set of only seven pieces (due to leaving one out) did not appear to capture all of the statistical regularities of Schenkerian analysis. Our corpus is the largest available and we are actively engaged in increasing its size. We are gathering analyses from music journals, textbooks, and Schenker's own published
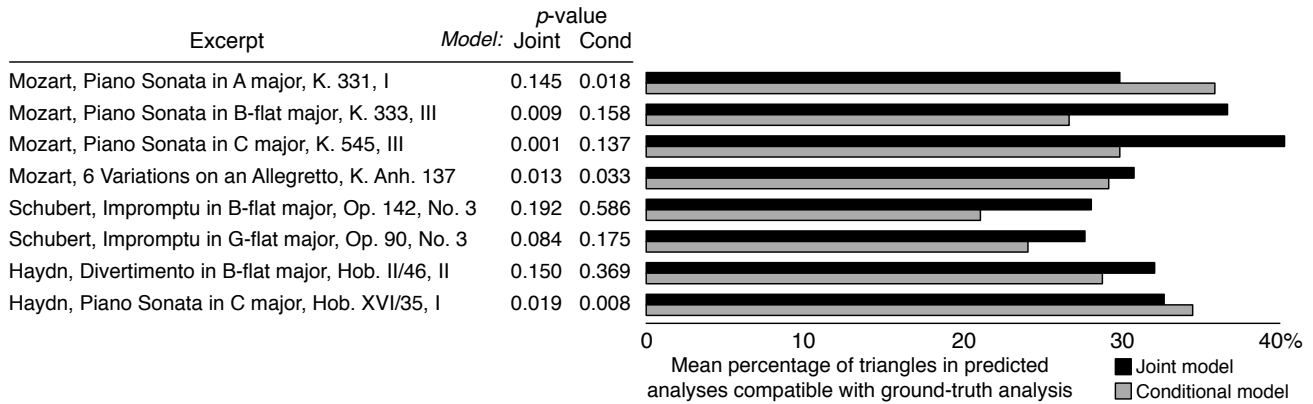
Excerpt	*Model:*	*p*-value Joint	Cond
Mozart, Piano Sonata in A major, K. 331, I		0.145	0.018
Mozart, Piano Sonata in B-flat major, K. 333, III		0.009	0.158
Mozart, Piano Sonata in C major, K. 545, III		0.001	0.137
Mozart, 6 Variations on an Allegretto, K. Anh. 137		0.013	0.033
Schubert, Impromptu in B-flat major, Op. 142, No. 3		0.192	0.586
Schubert, Impromptu in G-flat major, Op. 90, No. 3		0.084	0.175
Haydn, Divertimento in B-flat major, Hob. II/46, II		0.150	0.369
Haydn, Piano Sonata in C major, Hob. XVI/35, I		0.019	0.008

Mean percentage of triangles in predicted analyses compatible with ground-truth analysis. Joint model / Conditional model.

Figure 4. Leave-one-out cross-validation results for each of the eight excerpts in the corpus.

works. Third, we cannot overlook the possibility that OPT-MOP cannot produce a good analysis due to the model making incorrect assumptions or being too simplistic. Along with gathering more data, we are also working to improve our model of the analysis procedure.

7. CONCLUSIONS

Our work shows that actual Schenkerian analyses have statistical regularities that can be represented, discovered, and reproduced. We have shown statistically significant regularities in a data set of Schenkerian analyses and illustrated how those regularities may be exploited to design an algorithm for automatic analysis. Our experiment in ranking MOPs illustrates that assuming independence among the triangles comprising a MOP results in a satisfactory approximation to the joint probability of all the triangles. The probabilities of individual triangles in a MOP may be defined in numerous ways; in the future, we plan on collecting more contextual information surrounding prolongations, such as metrical positioning and harmonic information, and using these features to derive better probabilities over triangles.

8. REFERENCES

[1] Matthew Brown, Douglas Dempster, and Dave Headlam. The ♯IV(♭V) hypothesis: Testing the limits of Schenker's theory of tonality. *Music Theory Spectrum*, 19(2):155–183, 1997.

[2] Allen Forte and Steven E. Gilbert. *Introduction to Schenkerian Analysis*. W. W. Norton and Company, New York, 1982.

[3] R. E. Frankel, S. J. Rosenschein, and S. W. Smoliar. Schenker's theory of tonal music—its explication through computational processes. *International Journal of Man-Machine Studies*, 10(2):121–138, 1978.

[4] Robert E. Frankel, Stanley J. Rosenschein, and Stephen W. Smoliar. A LISP-based system for the study of Schenkerian analysis. *Computers and the Humanities*, 10(1):21–32, 1976.

[5] Édouard Gilbert and Darrell Conklin. A probabilistic context-free grammar for melodic reduction. In *Proceedings of the International Workshop on Artificial Intelligence and Music, 20th International Joint Conference on Artificial Intelligence*, pages 83–94, Hyderabad, India, 2007.

[6] Masatoshi Hamanaka and Satoshi Tojo. Interactive GTTM analyzer. In *Proceedings of the 10th International Society for Music Information Retrieval Conference*, pages 291–296, 2009.

[7] Phillip B. Kirlin and Paul E. Utgoff. A framework for automated Schenkerian analysis. In *Proceedings of the Ninth International Conference on Music Information Retrieval*, pages 363–368, 2008.

[8] Fred Lerdahl and Ray Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, Cambridge, Massachusetts, 1983.

[9] Alan Marsden. Automatic derivation of musical structure: A tool for research on Schenkerian analysis. In *Proceedings of the Eighth International Conference on Music Information Retrieval*, pages 55–58, 2007.

[10] Alan Marsden. Schenkerian analysis by computer: A proof of concept. *Journal of New Music Research*, 39(3):269–289, 2010.

[11] Panayotis Mavromatis and Matthew Brown. Parsing context-free grammars for music: A computational model of Schenkerian analysis. In *Proceedings of the 8th International Conference on Music Perception & Cognition*, pages 414–415, 2004.

[12] John Rahn. Logic, set theory, music theory. *College Music Symposium*, 19(1):114–127, 1979.

[13] Heinrich Schenker. *Der Freie Satz*. Universal Edition, Vienna, 1935. Published in English as *Free Composition*, translated and edited by E. Oster, Longman, 1979.

[14] Stephen W. Smoliar. A computer aid for Schenkerian analysis. *Computer Music Journal*, 2(4):41–59, 1980.

[15] Jason Yust. *Formal Models of Prolongation*. PhD thesis, University of Washington, 2006.