DISCOVERY OF CONTRAPUNTAL PATTERNS

Darrell Conklin
Department of Computer Science and AI
Universidad del País Vasco
San Sebastián, Spain
IKERBASQUE, Basque Foundation for Science
Bilbao, Spain
darrell_conklin@ehu.es

Mathieu Bergeron
CIRMMT
McGill University
Montreal, Canada
bergeron.mcgill@gmail.com

ABSTRACT

This paper develops and applies a new method for the discovery of polyphonic patterns. The method supports the representation of abstract relations that are formed between notes that overlap in time without being simultaneous. Such relations are central to understanding species counterpoint. The method consists of an application of the vertical viewpoint technique, which relies on a vertical slicing of the musical score. It is applied to two-voice contrapuntal textures extracted from the Bach chorale harmonizations. Results show that the new method is powerful enough to represent and discover distinctive modules of species counterpoint, including remarkably the suspension principle of fourth species counterpoint. In addition, by focusing on two voices in particular and setting them against all other possible voice pairs, the method can elicit patterns that illustrate well the unique treatment of the voices under investigation, e.g. the inner and outer voices. The results are promising and indicate that the method is suitable for computational musicology research.

1. INTRODUCTION

Polyphonic music presents challenges for music information retrieval, and the representation and discovery of patterns that recur in a polyphonic corpus remains an open and interesting problem. The discovery of monophonic patterns in music could be considered a mature area, with several powerful methods proposed and effectively applied to this task.

In polyphonic music, monophonic patterns can naturally be discovered using such methods provided that voice separation is first performed [12]. However, the discovery of polyphonic patterns – those containing two or more overlapping voices – is largely an open problem and there remain few approaches in the literature. The difficulty of this problem arises primarily from the presence of temporal relations between notes that occur in different voices, without being simultaneous. These relations are central to the understanding of counterpoint [9] and occur in all but the simplest note against note (first species) counterpoint. The discovery of polyphonic patterns is a new area with few existing approaches.

A difficulty with polyphony is that most polyphonic music cannot be easily separated into a regular number of voices [2]. In piano music, for example, voices can appear and disappear throughout a piece, and it would be mistaken to parse the music into a persistent number of voices. This paper uses the voiced Bach chorale harmonizations, and later some discussion is provided regarding the application of the method to unvoiced polyphony.

The simplest approach to find patterns in a polyphonic corpus is based on the conversion into chord symbol sequences, followed by the application of monophonic pattern discovery algorithms [1, 5]. However, accurate automatic chord labelling and segmentation is an unsolved problem, and the preparation of large corpora is difficult to achieve by hand.

Vertical approaches [4] are applied to voiced polyphonic textures first converted to a homophonic form. This is done by fully expanding a polyphonic score by adding artificial notes whenever a new onset time is encountered. This expanded score is then sliced, features computed for each slice in the sequence, and the resulting sequence mined using a monophonic pattern discovery approach. A similar method using bit-string approaches has been described by Chiu et al. [3]. The problem with these vertical approaches is in the lack of flexible temporal relations between components of a slice, which must all have the same onset time.

Geometric approaches [10] can be applied to voiced or unvoiced polyphony and can reveal patterns not contiguous on the music surface. The geometric approach however has two drawbacks for discovering contrapuntal patterns. Since geometric patterns must conserve exact inter-onset intervals in the time dimension, they cannot represent general local temporal relations among pattern components. Furthermore the expensive time complexity of the method leads to intractability for large pieces or corpora.

Polyphonic patterns are ideally represented as relational networks. Nodes represent events and edges represent relations between events. For example, consider the score fragment and small relational network presented in Figure 1(i,ii). This pattern matches four notes (a, b, c, d) in the following relations: a m7 dissonance (a, c) formed by the bottom voice (c) approached by a downwards leap, resolving to a M3 consonance (b, d) by a stepwise motion down in the top voice (b), and a leap up in the bottom voice (d). In addition, an asterisk $*$ indicates that (c) is preceded by a note that the pattern does not represent explicitly, but rather implicitly via the notion of melodic motion. Note that while (a, b) and (c, d) follow each other in the same voice, critically important in this pattern are the different temporal relations between (a, c) (c starting while a is sounding) and (b, d) (they start together).

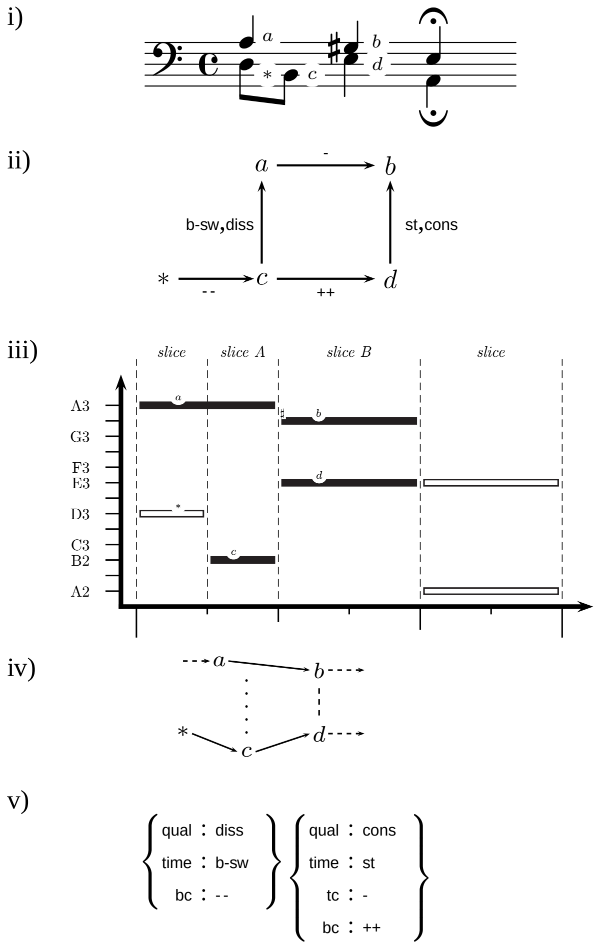


Figure 1. Two-voice counterpoint pattern: i) instance in final bar of BWV 257; ii) relational pattern; iii) piano-roll notation of instance showing the slicing process; iv) a schematic representation of the pattern; v) an equivalent feature set pattern.

Figure 1(iv) shows a schematic representation that we have developed as a visually appealing notation for a full relational network over a limited number of relations. Horizontal lines indicate voicing. The placement of vertical lines indicates temporal relations, and their form indicates consonance (dashed) or dissonance (dotted) between the temporally related notes. The slope of horizontal lines is proportional to the melodic contour it indicates (i.e. a steeper slope indicates a melodic leap). When the contour is not defined by the pattern, the horizontal line is dashed. Again, an asterisk $*$ indicates a note that must implicitly be present for the pattern to match, but is not explicitly matched by a pattern component.

This paper is concerned with discovering patterns like the one presented in Figure 1 in a large corpus of polyphonic music. We consider a precise notion of polyphonic pattern, one that captures musical intervals between notes that overlap in time, including between notes that are not simultaneous. Our approach is computationally very efficient as it translates the relational discovery problem to a form where sequential pattern discovery can be applied.

The paper is structured as follows. First, the new $\mathcal{VVP}$ (Vertical ViewPoint) method for polyphonic pattern discovery will be described. Results will be highlighted and translated to schematic relational networks for clarity. Finally, the strengths and weaknesses of $\mathcal{VVP}$ are discussed.

2. METHOD

The vertical viewpoint approach as originally presented [4] applies to homophony and cannot express temporal relations that might hold between components of a vertical slice, as would be required for the pattern developed in Figure 1. However, the approach can be extended resulting in what we call the $\mathcal{VVP}$ formalism. This is done by a) adding the ability to handle local temporal relations within slices; and b) describing properties of slices using flexible sets of features.

The first step was to employ event *continuations*, which specially identify those events with onsets prior to the onset of the slice. This method was inspired by the approach of Dubnov et al. [7] who developed a real-time method for prediction of continuations for polyphonic improvisations. Figure 1(iii) illustrates slices where continuations are used to retain local temporal relations. For example, the slice labelled A contains two notes: B2 in the bottom voice (c) and a continued event A3 in the top voice (a).

The second step was to employ *feature sets* [6] to describe successive slices in a piece. In this paper, four features are used to describe slices:

time describing the temporal relations between the two voices, taking the values st (both voices start together), b-sw (bottom voice starts while the top voice is sounding) and t-sw (top voice starts while the bottom voice is sounding). These temporal relations are derived by inspecting the continuation features of events within slices;

qual of the harmonic interval (non-compound diatonic interval between lower and higher pitches in the slice), taking the values cons (intervals P1, m3, M3, P5, m6, M6) and diss (all other intervals);

bc, tc describing the melodic contour of either the top voice (tc) or bottom voice (bc), both taking the possible values leap up: ++; leap down: --; step up: +; step down: -; unison: =.

Figure 1(v) illustrates a $\mathcal{VVP}$ pattern, using the above features. The pattern is a sequence of two features sets which expresses exactly the relational network in Figure 1(ii). It matches the score fragment in Figure 1(i) and in general all pairs of contiguous slices that contain the features presented.

To apply this representation in pattern discovery, a corpus is sliced at every unique onset time, every slice is saturated with the above features, then a pattern discovery algorithm finds all patterns that are distinctive (with a score above a threshold) and maximally general [5]. The algorithm for finding these patterns is a depth-first search of a subsumption space, that iteratively refines patterns at each search node to make them more specific. This is described in more detail in [5].

Regarding the scoring of patterns, it is well-known that discovered patterns should not be ranked simply by their count, because a general pattern will naturally occur more frequently than a more specific one, regardless of the corpus. Patterns are scored by an odds ratio, dividing their corpus probability by their background probability:

$$\Delta(P) \stackrel{\text{def}}{=} \frac{p(P|\oplus)}{p(P|\ominus)}, \tag{1}$$

where P is a pattern and the probabilities $p(P|\oplus)$ and $p(P|\ominus)$ are computed as follows. The probability of pattern P in the corpus is $p(P|\oplus) = c^{\oplus}(P)/n^{\oplus}$, where $c^{\oplus}(P)$ is the total slice count of pattern P in the corpus and $n^{\oplus}$ is the number of slices in the corpus. To evaluate the background probability $p(P|\ominus)$, two distinct methods may be used: the *anticorpus* method, and the the *null model* method.

An *anticorpus* is a set of pieces that is set up specifically to contrast with the analysis corpus. When an explicit anticorpus is used, the background probability is simply the anticorpus probability of the pattern:

$$p(P|\ominus) \stackrel{\text{def}}{=} c^{\ominus}(P)/n^{\ominus}, \tag{2}$$

where $c^{\ominus}(P)$ is the total slice count of the pattern P in the anticorpus, and $n^{\ominus}$ is the number of slices in the anticorpus.

In cases where an explicit anticorpus is not used, the *null model* method may be employed instead to rank patterns. The null model probability of a pattern P comprised of feature sets $c_1, \ldots, c_n$ is defined as:

$$p(P|\ominus) \stackrel{\text{def}}{=} \prod_{i=1}^{n} \prod_{f \in c_i} c^{\oplus}(f)/n^{\oplus}, \tag{3}$$

where $c^{\oplus}(f)$ is the total count (number of slices) of the feature f in the corpus. It estimates how probable a pattern is by multiplying the probabilities of its constituent features as they are encountered in the corpus. For example, the b-sw and t-sw temporal relations might occur less frequently than the st temporal relation. The null model probability of a pattern containing either b-sw or t-sw would then be lower, allowing for such a pattern to be interesting even if it has a lower total count than a similar pattern with a st temporal relation.

3. RESULTS

This section presents the results of $\mathcal{VVP}$ on pairs of voices extracted from 185 J.S. Bach chorale harmonizations in Humdrum format. Importantly, this format has diatonic spelling for all pitches, facilitating the computation of consonance and dissonance relations. Voices were extracted and recombined using the Humdrum toolset, and the appropriate filtering applied to insure well-formed slices (e.g. removing null lines to avoid slices where no event starts). All 6 possible ordered voice pairs were extracted from each piece: soprano-alto (SA), soprano-tenor (ST), soprano-bass (SB), alto-tenor (AT), alto-bass (AB), and tenor-bass (TB). This provides a total of $185 \times 6 = 1110$ possible two-voice pairs.

Two experiments were performed: one where an anticorpus was used to reveal the most interesting patterns and one where a null model was used. In both experiments, discovered pattern sets were filtered to retain only those patterns that contained at least a temporal (time) and harmonic interval quality feature (qual) in all components of a pattern. The method was calibrated to discover patterns with a score of at least 3 (Equation 1): with 3 times higher probability in the corpus as compared to the anticorpus or null model background probabilities (this calibration has worked well for other pattern discovery experiments in music [5]). In addition, only patterns with a total count of at least 100 were considered. This is a simple way to isolate patterns that are deemed to be reasonably frequent (the suspension pattern reported in Section 3.2, for example, has a total count of 450).

3.1 Anticorpus

In this experiment, all 6 voice pairs (SA, ST, SB, AT, AB, TB) were in turn used as the corpus, with the remaining 5 as the anticorpus. Thus each corpus comprises 185 two-voice polyphonic extracts, and each anticorpus $185 \times 5 = 925$ extracts. Patterns are ranked by using the anticorpus method (Equation 2) to compute background probabilities. Results for all experiments are concatenated, sorted from high to low $\Delta(P)$ (Equation 1), and the most distinctive 3 patterns each from the AT and SB corpora were retained (Table 1). The inner voices (AT) and outer voices (SB) were selected for illustration as they were previously discussed in the literature [8], but the method returns results for every voice pair. Below each pattern, Table 1 shows its total count and score as a number pair. The three patterns in each group are sorted by decreasing score $\Delta(P)$: note how a pattern can have higher score despite having a lower total count.

The first pattern of Table 1 is roughly 5 times more likely to occur between SB (outer voices) than any other voice pair. As mentioned by Huron [8], outer voices are more perceptually salient. They are hence expected to exhibit a good contrapuntal quality. Not surprinsingly, the patterns discovered by $\mathcal{VVP}$ as characteristic of SB are all consistent with counterpoint. For example, the first and third patterns introduce a dissonance over a b-sw temporal

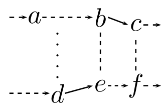
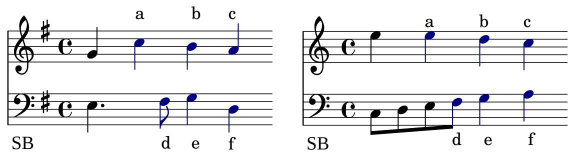
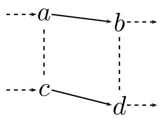
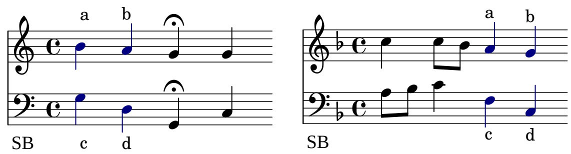
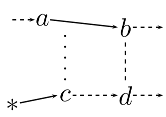
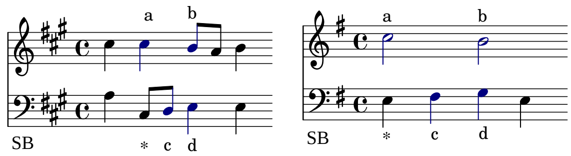
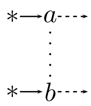
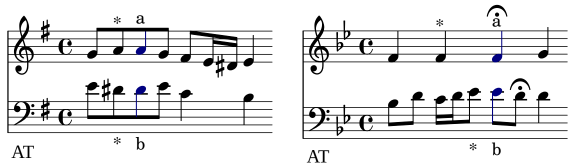
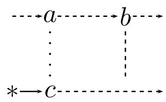
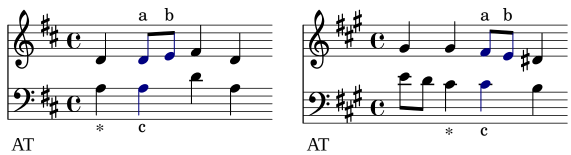
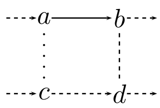
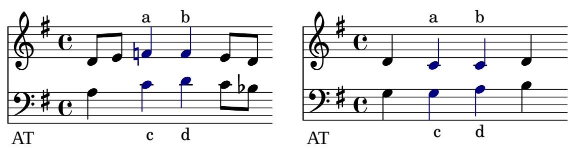
Pattern	Schema	Examples
{qual : diss, time : b-sw} {qual : cons, time : st, bc : +} {qual : cons, time : st, tc : -} (157, 4.88)		
{qual : cons, time : st} {qual : cons, time : st, tc : -, bc : --} (427, 4.78)		
{qual : diss, time : b-sw, bc : +} {qual : cons, time : st, tc : -} (271, 4.22)		
{qual : diss, time : st, tc : =, bc : =} (277, 7.08)		
{qual : diss, time : st, bc : =} {qual : cons, time : t-sw} (203, 6.89)		
{qual : diss, time : st} {qual : cons, time : st, tc : =} (533, 5.48)		

Table 1. Patterns discovered by computing background probabilities using an anticorpus. Top: distinctive of SB; bottom: distinctive of AT.

relation on a weak beat and resolve it straight away through stepwise melodic motion.

By opposition, one can expect the inner voices (AT), which are less perceptually salient, to be used more freely to create harmonies, with less considerations for the presence of dissonance. The last three patterns of Table 1 all characterize the inner voices and are consistent with this idea. All patterns introduce a dissonance over a `st` temporal relation. Pattern 5 clearly demonstrates this: a dissonances of a perfect fourth, usually forbidden between outer voices, here freely occurs in both examples.

In addition to the results presented in Table 1, many of the most interesting patterns discovered by $\mathcal{VVP}$ refer to either SB or AT pairs (data not shown). This suggests that the inner and outer voices are the two voice pairs that are the easiest to characterize, possibly because they differ the most from other voice pairs in terms of their contrapuntal quality.

3.2 Null model

In this experiment, we consider all voice pairs together (a corpus of 1110 two-voice pairs). Patterns are ranked by using a null model (Equation 3) to compute background probabilities. The seven most distinctive patterns are presented in Table 2. Again, the total count and score of each pattern is shown as a number pair.

The first pattern exhibits consecutive leaps. On its own, a leap is less likely than a step. Similarly, a sequence of leaps will be deemed unlikely by the null model. However, the first pattern contains such a sequence and does occur significantly in the chorales: 11 times more than expected. This can be interpreted as a composition rule: a sequence of leaps is acceptable, given that it occurs over two consonances, over two b-sw temporal relations, and changes contour. Note how the second instance of pattern 1 presents a rhythmic augmentation: the abstract pattern makes no constraints on rhythmic features.

The second, fourth and fifth pattern evoke second species counterpoint, while the sixth evokes first species. Pattern 3 evokes third species counterpoint, with a four note against one texture, alternating with dissonance and consonance. In one instance the bass line is rising: in the other it falls. In pattern 2, similarly, the direction of motion of the tenor line is inverted between the two instances shown. Again, this illustrates the abstraction power of the feature set pattern representation.

Pattern 7 is characteristic of the fourth species, with voices overlapping via successive b-sw and t-sw temporal relations. Remarkably, this is a pattern precisely describing the suspension pattern of fourth species counterpoint, including the melodic contour of a step down for the resolution of the dissonance [9]. The two instances presented are examples of the 4-3 and 7-6 suspension. Note how the high level of abstraction of this pattern (refraining from specifying exact intervals, and lower voice movement for the suspension resolution) is necessary to represent the concept of a suspension.

4. DISCUSSION

This paper presented the $\mathcal{VVP}$ method for the discovery of relational patterns in two-voice counterpoint. The method is based on a monophonic pattern discovery algorithm [5], and extends earlier results [4] through the use of continuations of events across slices. The $\mathcal{VVP}$ approach is fast as it transforms a relational data mining problem into a simpler sequential one, an example of representation change often employed in machine learning. Nevertheless, the representation is flexible due to the abstraction of slices by feature sets.

In the experiments presented here, the representation allows the concise expression of many contrapuntal patterns. This demonstrates that the approach is powerful enough to discover polyphonic patterns of theoretical significance. The patterns reported here are not particularly surprising, but their discovery is nonetheless promising: further exploration with the method could help discover significant patterns that are yet unknown to music theory. This exploration is however outside the scope of the current paper.

Similar patterns to the those presented here have been studied in the context of supervised learning [11], where the patterns were identified beforehand and the task was to learn rules explaining the structure of the patterns. However, only simple note against note counterpoint was studied.

The application of the method in this paper led to patterns that were very abstract, with highly divergent instances on the musical surface. It was shown, however, that such abstraction is necessary to capture the concept of a fourth species suspension pattern in its full generality. Furthermore, using only three simple background features (temporal relations, harmonic interval quality, and melodic contour), the method was able to discover the suspension in a corpus of Bach chorale harmonizations.

If less abstraction is desired, more background features can be added, for example those referring to melodic interval and durations of notes. For this study, a calibration of the method using the most basic relations of counterpoint was desired. This has been successful and future work will focus on further pattern representation aspects.

Though applied to two-voice textures here, the method easily extends to more voices. Further temporal relations would need to be added to accommodate the additional temporal relations formed; for example for pattern discovery in 3 voices (e.g., SAT), three overlap relations (SA, ST, AT) and three corresponding simultaneity relations would be needed: in general a number of such features quadratic in the number of notes in a slice. In addition, three different harmonic interval relations would be necessary, or alternatively the notion of harmonic interval quality could somehow be extended to triads.

A limitation of this approach, as presented here, is that it applies to voiced polyphony, or corpora where voices have already been extracted from an unvoiced polyphonic texture. There are several possible ways to apply the method to raw unvoiced textures. Temporal relations between slice components may simply be discarded, and any “static” viewpoint (such as chord type, mode) of a slice may be used. Alternatively, the presence of a temporal relation between slice components can be considered but with voice leading relations (e.g., melodic contour) omitted. Finally, if both temporal and voice-leading relations are desired, one might replace the notions of top voice and bottom voice with the highest note and lowest note in the slice, reminiscent of the simple skyline approach to voice segregation. With feature sets, one could even include melodic relationships between second highest pitches (if present), third highest pitches (if present), lowest (possible bass line) and so on. The presence or absence of such features would also indicate texture density, hence potentially expressing patterns occurring over changes of textures.

5. ACKNOWLEDGEMENTS

Darrell Conklin is supported by IKERBASQUE, Basque Foundation for Science, Bilbao, Spain. Mathieu Bergeron is supported by Le Fonds québécois de la recherche sur la nature et les technologies, Québec, Canada.

6. REFERENCES

[1] A. Anglade and S. Dixon. Characterisation of harmony with inductive logic programming. In *International Conference on Music Information Retrieval (ISMIR)*, pages 63–68, Philadelphia, USA, 2008.

[2] E. Cambouropoulos. Voice and stream: Perceptual and computational modeling of voice separation. *Music Perception*, 26(1):75–94, 2008.

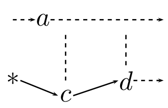
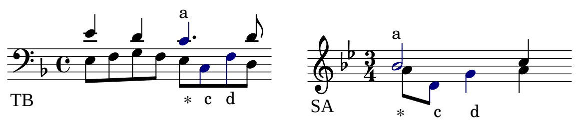
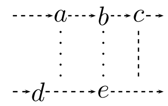
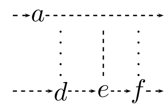
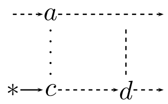
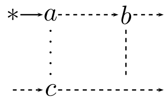
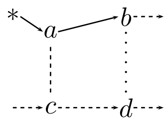
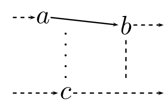
Pattern	Schema	Examples
{qual : cons, time : b-sw, bc : --} {qual : cons, time : b-sw, bc : ++} (152, 11.69)		
{qual : diss, time : t-sw} {qual : diss, time : st} {qual : cons, time : t-sw} (158, 11.68)		
{qual : diss, time : b-sw} {qual : cons, time : b-sw} {qual : diss, time : b-sw} (154, 10.11)		
{qual : diss, time : st, bc : =} {qual : cons, time : b-sw} (1128, 9.69)		
{qual : diss, time : st, tc : =} {qual : cons, time : t-sw} (804, 9.00)		
{qual : cons, time : st, tc : --} {qual : diss, time : st, tc : ++} (116, 8.89)		
{qual : diss, time : b-sw} {qual : cons, time : t-sw, tc : -} (450, 7.73)		

Table 2. Patterns discovered by computing background probabilities using a null model.

[3] S-C. Chiu, M-K. Shan, J-L. Huang, and H-F. Li. Mining polyphonic repeating patterns from music data using bit-string based approaches. In *ICME 2009*, pages 1170–1173, 2009.

[4] D. Conklin. Representation and discovery of vertical patterns in music. In *Music and Artificial Intelligence: Lecture Notes in Artificial Intelligence*, number 2445, pages 32–42. Springer-Verlag, 2002.

[5] D. Conklin. Discovery of distinctive patterns in music. *Intelligent Data Analysis*, 14(5), 2010. To appear.

[6] D. Conklin and M. Bergeron. Representation and discovery of feature set patterns in music. *Computer Music Journal*, 32(1):60–70, 2008.

[7] S. Dubnov, G. Assayag, O. Lartillot, and G. Bejerano. Using machine-learning methods for musical style modeling. *IEEE Computer*, 36(10):73–80, 2003.

[8] D. Huron. Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception*, 6(4):361–382, 1989.

[9] K. Kennan. *Counterpoint*. Prentice-Hall, 1959.

[10] D. Meredith, K. Lemström, and G. A. Wiggins. Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music. *Journal of New Music Research*, 31(4):321–345, 2002.

[11] U. Pompe, I. Kononenko, and T. Makše. An application of ILP in a musical database: Learning to compose the two-voice counterpoint. In *Proceedings of the MLnet Familiarization Workshop on Data Mining with Inductive Logic Programing*, pages 1–11, Heraklion, Crete, 1996.

[12] P. E. Utgoff and P. B. Kirlin. Detecting motives and recurring patterns in polyphonic music. In *ICMC*, pages 487–494, New Orleans, Lousianna, USA., 2006.