SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type	conference contribution
Title	Sequential association rules in atonal music
Author(s)	A. Honingh, T. Weyde, D. Conklin
Faculty	FNWI: Institute for Logic, Language and Computation (ILLC)
Year	2009

FULL BIBLIOGRAPHIC DETAILS:
http://hdl.handle.net/11245/1.324425

Sequential Association Rules in Atonal Music

Aline Honingh*, Tillman Weyde and Darrell Conklin

Music Informatics research group
Department of Computing
City University London

Abstract. This paper describes a preliminary study on the structure of atonal music. In the same way as sequential association rules of chords can be found in tonal music, sequential association rules of pitch class set categories can be found in atonal music. It has been noted before that certain pitch class sets can be grouped into 6 different categories [10]. In this paper we calculate those categories in a different way and show that virtually all possible pitch class sets can be grouped into these categories. Each piece in a corpus of atonal music was segmented at the bar level and of each segment it was calculated to which category it belongs. The percentages of occurrence of the different categories in the corpus were tabulated, and it turns out that these statistics may be useful for distinguishing tonal from atonal music. Furthermore, sequential association rules were sought within the sequence of categories. The category transition matrix shows how many times it happens that one specific category is followed by another. The statistical significance of each progression can be calculated, and we present the significant progressions as sequential association rules for atonal music.

keywords: pitch class set categories, atonal music, sequential association rules, similarity measures.

1 Introduction

A typical structure can usually be revealed in tonal music, when it is analyzed harmonically. The chord progressions like the ones shown in Table 1 show some general rules that can often be found in Western tonal music. Atonal music, on the other hand, is not structured around a tonal center like tonal music. Therefore, for atonal music, a progression table like this is impossible. Pitch class set theory can be used to analyze atonal music and more analysis theories have been proposed to analyze atonal music [7]. However, no analogy to chord progression in tonal music has been proposed. In this paper, we will ask ourselves the question whether any kind of progression rules for atonal music can be found which could reveal partly the structure of atonal music.

* *Corresponding author address:*
Institute for Logic, Language and Computation
University of Amsterdam, The Netherlands
e-mail: A.K.Honingh@uva.nl

Table 1. Chord progression in major mode, taken from [9].

Chord	is followed by	sometimes by	less often by
I	IV, V	VI	II, III
II	V	IV, VI	I,III
III	VI	IV	I, II, V
IV	V	I, II	III, VI
V	I	VI, IV	III, II
VI	II, V	III, IV	I
VII	III	I	

1.1 Atonal Music and Pitch Class Set Theory

A distinction is often made between "free" atonal music and twelve tone or serial music. Twelve tone music differs from free atonal music in two important ways: all 12 pitch classes are used and ordered. In this paper, when we speak about atonal music, we mean both free atonal music and serial music.

For the analysis of atonal music, pitch class set theory has been developed. Pitch class set theory has been described in 1973 by Alan Forte [4]. A pitch class is a number between 0 and 11 and is an abstraction of a musical note. All 12 pitch classes represent the semitones from one octave. Collections of pitch class sets (harmonic or melodic) can be analyzed according to pitch class set theory. Forte [4] assumed two types of equivalence (besides the octave equivalence and enharmonic equivalence that belong to pitch classes) related to collections, namely transpositional equivalence and inversional equivalence. Furthermore, the term 'set' covers permutation equivalence and cardinality equivalence. For example, the set $\{0, 4, 7\}$ represents all the chords/melodies that are composed of these three pitch classes (including repetitions). Without these equivalence classes, the number of possible pitch class sets would be huge. But taking these equivalence relations into account, the list of possible pitch class sets are more limited and each set in the list can be characterized by the so-called prime form of the pitch class set (see e.g. [4] for more information). An other way of describing a pitch class set is to characterize it by its intervallic content. The interval-class vector or IcV is an array that expresses the intervallic content of a pitch class set. Since in pitch class set theory an interval is equal to its inverse, an IcV consists of six numbers instead of twelve, with each number representing the number of times an interval class appears in the set. For example, the pitch class set $\{0, 4, 7\}$ has interval class vector [0 0 1 1 1 0] since it consists of 1 'minor third or major sixth', 1 'major third or minor sixth', and 1 'perfect fourth or perfect fifth'. An IcV represents a pitch class set together with all its transformations according to the above mentioned equivalence classes.

Although the list of different pitch class sets according to Forte may be limited, it still consists of 351 sets (that is the list of different prime forms, [4]), and therefore similarity measures are sometimes useful.

2 Pitch Class Set Categories in Atonal Music

Many similarity measures have been developed for pitch class sets, for example Isaacson's IcVSIM [6], Forte's R_n relations [4], Morris' SIM [8], Rahn's MEMB [11], Rogers' cosθ [12] and Scott and Isaacson's Angle [13]. Many of these similarity measures are based on the interval-class vector (IcV), i.e. those measures compare two different IcV's and output a value that characterizes their similarity.

At first sight, these similarity measures seem to be not so much related to each other. They differ in range, intention, way of calculation and more [10]. However, Quinn argues that those similarity measures have actually a lot in common: they tend to group the IcV's in six different categories, each of which can be said to correspond to a cycle of one of the six interval classes [10]. A cycle of interval classes can be thought of in the following way. A cycle of the interval 1 will read: 0,1,2,3,4, ... A cycle of the interval 2 will read: 0,2,4,6,.... A cycle of the interval 3 will read: 0,3,6,9,... , and so on. Using a cluster analysis, Quinn groups the tetrachords and pentachords in six categories according to several different similarity measures. He identifies for each category a prototype. If a certain pitch class set is grouped into a certain category, this pitch class set is similar to the prototype of that category, according to the similarity measure used. The set $\{0, 1, 2, 3, 4\}$ (IcV=[4 3 2 1 0 0]) is the prototype of the Interval Category 1 (IC1) in the pentachord classification, the set $\{0, 2, 4, 6, 8\}$ (IcV=[1 3 1 2 2 1]) the prototype of IC2, and so on. The cycles of IC's that have periodicities that are less than the cardinality of their class (for example, pitch class 4 has a periodicity of 3: {0,4,8}) are extended in the way described by Hanson [5]: the cycle is shifted to pitch class 1 and continued from there. For example, the IC-6 cycle proceeds $\{0, 6, 1, 7, 2, 8...\}$ and the IC-4 cycle proceeds $\{0, 4, 8, 1, 5, 9, 2, ...\}$. Thus for every cardinality, a separate prototype characterizes the category. For example, category IC4 has prototype $\{0, 4\}$ for sets of cardinality 2, prototype $\{0, 4, 8\}$ for set of cardinality 3 and so on. Tables 2 and 3 give an overview of the prototypes of pitch class set categories. Prototypes have been listed for sets from 2 to 10 notes. Pitch class sets with less than 2 notes or more than 10 notes do not make sense. One pitch class set of cardinality 1 exists, $\{0\}$, with interval vector [0 0 0 0 0 0] and it belongs equally to every category. The same is true for cardinality 11: only one prime form pitch class set exists: $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ with interval vector [10 10 10 10 10 5] and belongs to every category equally. The pitch class set of cardinality 12 contains all possible pitch classes.

Although the general classification into six categories is clear from the cluster analysis by Quinn [10], some differences can still be found between classifications with respect to different similarity measures. Comparing the clusters that are obtained from the cluster analysis by Quinn on IcVSIM [6] and SATSIM [1], it appears that two sets, $\{0, 1, 2, 5, 7\}$ and $\{0, 1, 3, 6, 8\}$, that are categorized by IcVSIM as IC5 are categorized by SATSIM as IC6 (see [10]). More differences exist in the classifications when a comparison is made with more similarity measures. Aiming to group the pitch class sets uniformly, in this paper a slightly different approach will be used to classify the pitch class sets. We have used the

Table 2. Prototypes expressed in pitch class sets for the six categories

	prototypes (pc sets)
IC1	$\{0,1\}$, $\{0,1,2\}$, $\{0,1,2,3\}$, etc.
IC2	$\{0,2\}$, $\{0,2,4\}$, $\{0,2,4,6\}$, etc.
IC3	$\{0,3\}$, $\{0,3,6\}$, $\{0,3,6,9\}$, etc.
IC4	$\{0,4\}$, $\{0,4,8\}$, $\{0,1,4,8\}$, etc.
IC5	$\{0,7\}$, $\{0,2,7\}$, $\{0,2,5,7\}$, etc.
IC6	$\{0,6\}$, $\{0,1,6\}$, $\{0,1,6,7\}$, etc.

Table 3. Prototypes expressed in interval class vectors for the corresponding classes of different cardinality.

	prototypes (IcV)					
	IC1	IC2	IC3	IC4	IC5	IC6
duochord classes	[1 0 0 0 0 0]	[0 1 0 0 0 0]	[0 0 1 0 0 0]	[0 0 0 1 0 0]	[0 0 0 0 1 0]	[0 0 0 0 0 1]
trichord classes	[2 1 0 0 0 0]	[0 2 0 2 0 0]	[0 0 2 0 0 1]	[0 0 0 3 0 0]	[0 1 0 0 2 0]	[1 0 0 0 1 1]
tetrachord classes	[3 2 1 0 0 0]	[0 3 0 2 0 1]	[0 0 4 0 0 2]	[1 0 1 3 1 0]	[0 2 1 0 3 0]	[2 0 0 0 2 2]
pentachord classes	[4 3 2 1 0 0]	[1 3 1 2 2 1]	[1 1 4 1 1 2]	[2 0 2 4 2 0]	[0 3 2 1 4 0]	[3 1 0 1 3 2]
hexachord classes	[5 4 3 2 1 0]	[0 6 0 6 0 3]	[2 2 5 2 2 2]	[3 0 3 6 3 0]	[1 4 3 2 5 0]	[4 2 0 2 4 3]
heptachord classes	[6 5 4 3 2 1]	[2 6 2 6 2 3]	[3 3 6 3 3 3]	[4 2 4 6 4 1]	[2 5 4 3 6 1]	[5 3 2 3 5 3]
octachord classes	[7 6 5 4 4 2]	[4 7 4 6 4 3]	[4 4 8 4 4 4]	[5 4 5 7 5 2]	[4 6 5 4 7 2]	[6 4 4 4 6 4]
nonachord classes	[8 7 6 6 6 3]	[6 8 6 7 6 3]	[6 6 8 6 6 4]	[6 6 6 9 6 3]	[6 7 6 6 8 3]	[7 6 6 6 7 4]
decachord classes	[9 8 8 8 8 4]	[8 9 8 8 8 4]	[8 8 9 8 8 4]	[8 8 8 9 8 4]	[8 8 8 8 9 4]	[8 8 8 8 8 5]

prototypes themselves to classify pitch class sets into the aforementioned six categories by using the chosen similarity measure to calculate to which prototype a pitch class set is closest. When doing this, the categorization of pentachords according to the aforementioned similarity measures IcVSIM and SATSIM are identical, so Quinn's [10] claim about the six categories could be made even stronger. Even more similarity measures could be compared in this respect. We have compared the measures IcVSIM [6], SATSIM [1], ASIM [8] and cosθ [12], and found they all come up with the same classification for the duochords, pentachords, heptachords, octachords, nonachords and decachords, and the classifications for the trichords, tetrachords and hexachords differ at most by 3 pitch class sets. This shows that similarity measures are not too different in this respect, they agree on the classification in the six categories as we find a very high overlap.

We will base our choice of which similarity measure to use, on the ambiguity it produces. It turns out that Rogers' cosθ produces the least ambiguity: when

using it to calculate the category of a pitch class set, it outputs virtually always only one category.

3 Sequential Association Rules

Each category can be seen to as having a particular character resulting from the intervals that appear most frequently. Category 1 (see Table 2 for the prototypes) consists of all semitones and is the category of the chromatic scale. Category 2 is the category of the whole-tones or whole-tone scale. Category 3 is the category of the diminished triads or diminished scale. Category 4 is the category of the augmented triads or augmented scale. Category 5 is the category of the diatonic scale. Category 6 is the category of the tritones or D-type all-combinatorial hexachord (see [5]).

As we have shown above, four similarity measures group the pitch class sets into the same categories. Since similarity plays a role in the analysis of music, this might suggest that those categories play a structural role in atonal music. In this paper we will try to discover sequential association rules between those categories in a corpus of atonal music such as to come up with a table of 'category progressions' for atonal music similar to that of Piston [9] for tonal music. A sequential association rule is a progression $a \to b$, where the probability $p(b|a)$ is higher than chance level, meaning that category b tends to follow category a more often than expected [2].

3.1 The Method

The method has been implemented in Java, using parts of the Musitech Framework [14], and operates on MIDI data. The MIDI file is segmented on the bar level, as a first step to investigate the raw regularities that occur on this level. The pitches from each bar form a pitch class set. From each pitch class set, the interval class vector can be calculated after which the category it belongs to can be calculated. Using Rogers' $\cos\theta$ as similarity measure we calculate the similarity to all prototypes of the required cardinality. The prototype to which the set is most similar, represents the category to which the set belongs. If the the pitch class set that is constructed from a bar contains less than 2 or more than 10 different pitch classes, the category is not calculated since this does not make any sense, as we explained in Sect. 2. Therefore, if a set (bar) contains more than 10 different pitch classes, the bar is divided into beats and the beats are treated as new pitch class sets. If a set contains less than 2 pitch classes, this set is added to the set that is constructed from the next bar.

First of all, the number of occurrences of all categories are counted, such that we get an overview of the piece in terms of the percentages of occurrence of the different categories. Furthermore, the instances of each progression from one category to another are counted.

A measure for the over-representation of a progression $a \rightarrow b$ is the 'lift'. This measure is taken from [2] and defined as follows:

$$\text{lift}(a \rightarrow b) = \frac{p(b|a)}{p(b)}, \tag{1}$$

where $p(x)$ denotes the probability of category x. The lift can be understood as the number of observed progressions divided by the number of expected progressions due to chance. If the lift is greater than 1, there is a positive correlation, if the lift is smaller than 1, there is a negative correlation.

4 Results

As described in the previous section, the occurrences of each category can be counted. It can be expected that different types of music will show a different occurrence rate for each category. To start with a tonal piece, for example, the distribution of categories of the fourth movement of Beethoven's ninth symphony is shown is Table 4. One can observe that category 5 dominates the whole piece.

Table 4. Distribution of categories of the fourth movement of Beethoven's ninth symphony.

category	number of occurrences	percentage of occurrence
1	102	11.40 %
2	69	7.71 %
3	78	8.72 %
4	89	9.94 %
5	552	61.68 %
6	5	0.56 %

This turns out to be quite typical for tonal music. In the previous section we have mentioned that each category can be seen as having a specific character and category 5 represents the diatonic scale. Therefore, it is not surprising that a piece of tonal music based on the diatonic scale is dominated by category 5.

For atonal music, we expect something different. We have run the program on atonal music of Schoenberg, Webern, Stravinsky and Boulez. The complete list of music is shown in Table 5. On average, the distribution as shown in Table 6 was found, using this corpus of atonal music. One can see that this distribution is totally different from Table 4 and as such this method might be useful in discrimination tasks. We can see that the music is not dominated anymore by category 5 but a much more equal distribution is present in atonal music.

A transition matrix can be made with our method (Table 7), listing how many times category i is followed by category j.

Table 5. The atonal music used in the method

composer	**piece**
Schoenberg	Pierrot Lunaire part 1, 5, 8, 10, 12, 14, 17, 21
Schoenberg	Piece for piano opus 33
Schoenberg	Six little piano pieces opus 19 part 2, 3, 4, 5, 6
Webern	Symphony opus 21 part 1
Webern	String Quartet opus 28
Boulez	Notations part 1
Boulez	Piano sonata no 3, part 2: "Texte"
Boulez	Piano sonata no 3, part 3: "Parenthese"
Stravinsky	in memoriam Dylan Thomas Dirge canons (prelude)

Table 6. Distribution of categories from music of Schoenberg, Webern, Stravinsky and Boulez

category	number of occurrences	percentage of occurrence	standard deviation
1	313	28.25 %	10.56 %
2	117	10.56 %	6.14 %
3	166	14.98 %	7.68 %
4	179	16.16 %	7.97 %
5	138	12.45 %	7.15 %
6	195	17.60 %	6.20 %

We have calculated the lift matrix as described in the previous section (Table 8) from which one can see which progressions have a positive relation and which have a negative relation.

To answer the question which progressions are meaningful, we have to perform a significance test. We would like to know which progressions have an occurrence rate that is significantly higher or lower than chance level. We use a chi-square test on the data of Table 7 to calculate which progressions cannot be explained by our null hypothesis: the probability of class j following class i does only depend on the overall number of j's in the music. We calculate the chi-square statistics for every progression separately by making a 2×2 contingency table (with fields $i \rightarrow j$, $i \rightarrow \neg j$, $\neg i \rightarrow j$, $\neg i \rightarrow \neg j$), and calculate the probability from the probability density function of the chi-square distribution with 1 degree of freedom (Table 9). If we take the significance level to be 5%, the progressions that are significantly meaningful are printed in boldface in Table 9.

Now that we can identify the meaningful progressions for our corpus of atonal music, we can make a table for categories analogue to Piston's table for chords. From the lift value in Table 8 can be seen whether a significant progression represents a positive or negative association. These significant rules can be found in Table 9 under the headings "is followed by" (positive association) and "less often by" (negative association). One can see that there is a tendency for categories

Table 7. The transition matrix

category	To 1	2	3	4	5	6
From 1	109	23	49	36	28	62
2	27	21	12	15	18	22
3	49	18	30	21	24	22
4	44	17	29	39	15	32
5	33	17	15	29	27	16
6	47	20	28	34	22	38

Table 8. The lift matrix

category	To 1	2	3	4	5	6
From 1	1.23	0.70	1.04	0.71	0.72	1.13
2	0.82	1.70	0.68	0.79	1.24	1.07
3	1.04	1.03	1.21	0.78	1.16	0.75
4	0.87	0.90	1.08	1.35	0.67	1.02
5	0.85	1.17	0.73	1.30	1.57	0.66
6	0.85	0.97	0.96	1.08	0.91	1.11

Table 9. The significance matrix of the results displayed in Table 7.

category	To 1	2	3	4	5	6
From 1	**0.001**	**0.020**	>0.5	**0.009**	**0.026**	0.111
2	0.150	**0.003**	0.097	0.288	0.179	>0.5
3	>0.5	>0.5	0.135	0.159	0.252	0.079
4	0.201	>0.5	>0.5	**0.008**	0.059	>0.5
5	0.163	0.438	0.104	**0.047**	**0.003**	**0.030**
6	0.167	>0.5	>0.5	0.345	1.222	0.254

to follow itself, so that large regions in the music are represented by just one category. This is in accordance with observations by Ericksson [3], who describes 7 categories similar to the ones described above and says that "it is often possible to show that one region [category] dominates an entire section of a piece". Besides these 'repetitions' of categories, one other progression can be identified to present a sequential association rule: the progression from 5 to 4, and four other progressions can be identified to present a negative association, sequential 'avoidance' rules: the progression from category 1 to 2, from 1 to 4, from 1 to 5, and from 5 to 6.

Table 10. Category progression in atonal music.

Category	is followed by	sometimes by	less often by
1	1	3,6	2,4,5
2	2	1,3,4,5,6	
3		1,2,3,4,5,6	
4	4	1,2,3,5,6	
5	4,5	1,2,3	6
6		1,2,3,4,5,6	

5 Concluding Remarks

Although this work serves as a preliminary study on sequential association rules in atonal music, some interesting things can be said. To sum up the results of this paper, we showed first of all that the 6 different pitch class categories described in [10], can be found in a different way by comparing all pitch class sets to certain prototypes according to a specific similarity measure. Four different similarity measures agree virtually always on the grouping of all possible pitch class sets into these 6 categories. Furthermore, the distribution of notes into these categories appears to be distinguishing between atonal and tonal music and could perhaps be used as a tool for this purpose. Finally, a number of sequential association rules have been found in a corpus of atonal music. A sequential association rule is a progression from category i to j that appears in the music significantly more often than one would expect due to chance. These progression rules may reveal a structure of atonal music that was not known before.

6 Acknowledgements

We wish to thank our colleague Mathieu Bergeron for presenting this paper at the second international conference of the Society for Mathematics and Computation in Music 2009.

References

1. Michael Buchler. Relative saturation of interval and set classes: A new model for understanding pcset complementation and resemblance. *Journal of Music Theory*, 45(2):263–343, 2001.
2. Darrell Conklin. Melodic analysis with segment classes. *Machine Learning*, 65(2-3):349–360, 2006.
3. Tore Ericksson. The ic max point structure, mm vectors and regions. *Journal of Music Theory*, 30(1):95–111, 1986.
4. Alan Forte. *The Structure of Atonal Music*. New Haven: Yale University Press, 1973.
5. Howard Hanson. *Harmonic Materials of Modern Music*. New York: Appleton-Century-Crofts, 1960.
6. Eric J. Isaacson. Similarity of interval-class content between pitch-class sets: the IcVSIM relation. *Journal of Music Theory*, 34:1–28, 1990.
7. Fred Lerdahl. Atonal prolongational structure. *Contemporary Music Review*, 4(1):65–87, 1989.
8. Robert Morris. A similarity index for pitch-class sets. *Perspectives of New Music*, 18:445–460, 1980.
9. Walter Piston and Mark DeVoto. *Harmony*. Victor Gollancz Ltd, revised and expanded edition, 1989.
10. Ian Quinn. Listening to similarity relations. *Perspectives of New Music*, 39:108–158, 2001.

11. John Rahn. Relating sets. *Perspectives of New Music*, 18:483–498, 1980.
12. David W. Rogers. A geometric approach to pcset similarity. *Perspectives of New Music*, 37(1):77–90, 1999.
13. Damon Scott and Eric J. Isaacson. The interval angle: A similarity measure for pitch-class sets. *Perspectives of New Music*, 36(2):107–142, 1998.
14. Tillman Weyde. Modelling cognitive and analytic musical structures in the MUSITECH framework. In *UCM 2005 5th Conference "Understanding and Creating Music", Caserta, November 2005*, pages 27–30, 2005.