

RHYTHM COMPLEXITY MEASURES: A COMPARISON OF MATHEMATICAL MODELS OF HUMAN PERCEPTION AND PERFORMANCE

Eric Thul
School of Computer Science
Schulich School of Music
McGill University, Montréal
ethul@cs.mcgill.ca

Godfried T. Toussaint
School of Computer Science
Schulich School of Music
McGill University, Montréal
godfried@cs.mcgill.ca

ABSTRACT

Thirty two measures of rhythm complexity are compared using three widely different rhythm data sets. Twenty-two of these measures have been investigated in a limited context in the past, and ten new measures are explored here. Some of these measures are mathematically inspired, some were designed to measure syncopation, some were intended to predict various measures of human performance, some are based on constructs from music theory, such as Pressing's cognitive complexity, and others are direct measures of different aspects of human performance, such as perceptual complexity, meter complexity, and performance complexity. In each data set the rhythms are ranked either according to increasing complexity using the judgements of human subjects, or using calculations with the computational models. Spearman rank correlation coefficients are computed between all pairs of rhythm rankings. Then phylogenetic trees are used to visualize and cluster the correlation coefficients. Among the many conclusions evident from the results, there are several observations common to all three data sets that are worthy of note. The syncopation measures form a tight cluster far from other clusters. The human performance measures fall in the same cluster as the syncopation measures. The complexity measures based on statistical properties of the inter-onset-interval histograms are poor predictors of syncopation or human performance complexity. Finally, this research suggests several open problems.

1 INTRODUCTION

Many music researchers consider rhythm to be the most important characteristic of music. Furthermore, one of the main features of rhythm is its complexity. Therefore measures of the complexity of a rhythm constitute key features useful for music pattern recognition and music information retrieval, as well as ethnomusicological analyses of world music [17, 18]. Since the notion of complexity is flexible, it is not surprising that in the past a variety of different measures of complexity has appeared in the literature. Areas

where such measures have been applied range from psychology, engineering, computer science and mathematics, to music theory. Given such a wide range of applicable fields, different techniques for measuring complexity have been developed. For example, one can analyze a rhythm's binary sequence representation, ask listeners to rate a rhythm's complexity, or ask musicians to perform a rhythm. Therefore, in our work, we include measures of information and coding complexity, performance complexity, and cognitive complexity. Furthermore, there are traditional concepts in music such as *syncopation* [10] which may also be considered as measures of rhythm complexity [7, 8].

With the exception of [7, 8], previous research on rhythm complexity has been limited to determining how good a feature it is for music pattern recognition, or how well it models human judgements of complexity [17, 18]. Moreover, for such studies researchers have used data (families of rhythms) that were generated artificially and randomly with some constraints. Here, we not only use a large group comprised of 32 measures of complexity that employ a wide variety of measurement techniques, but we also validate these measures against human judgements of perceptual, meter, and performance complexity using three diverse data sets.

2 COMPLEXITY MEASURES

One can broadly categorize the complexity measures used in this study into two distinct categories: human performance measures directly obtained from psychological experiments, and measures obtained from mathematical models of rhythm complexity. The human performance measures can be subdivided into three types: *perceptual* complexity, *meter* complexity, and *performance* complexity. Perceptual complexity is obtained by asking human subjects to judge complexity as they listen to rhythms. Meter complexity is obtained by measuring how well the human subjects are able to track the underlying *metric beat* of a rhythm. It is worth noting that some researchers, for example in music psychology [4], refer to the metric beat as the *pulse*.

Here we reserve the word *pulse* for the largest duration interval that evenly divides into all the inter-onset onsets (IOI) present in a family of rhythms. This is common terminology in ethnomusicology and music technology. Performance complexity measures pertain to how well the subjects can reproduce (execute, play-back) the rhythms, usually by tapping. The mathematical models can be subdivided into two main categories: those that are designed to measure *syncopation*, and those that are designed to measure *irregularity*. The irregularity measures can be divided into *statistical* and *minimum-weight-assignment* measures.

Due to lack of space, we cannot provide a detailed description of all the complexity measures tested. Thus we list the complexity measures with each corresponding essential reference in the literature for further information, along with a label in parentheses pertaining to the phylogenetic tree labels used in Figures 1, 2, and 3. Measures of syncopation are listed first. The Longuet-Higgins and Lee measure (*lhl*) [4, 14], along with Smith and Honing's version (*smith*) [19], take advantage of a metric hierarchy of weights [13] to calculate syncopation. A variation of Tossaint's metrical complexity (*metrical*) [21] and Keith's measure (*keith*) [10] also use this hierarchy to judge syncopation. The Weighted Note-to-Beat Distance (*wmbd*, *wmbd2*, *wmbd4*, *wmbd8*) [7] uses the distance from onsets to *metric beats* to gauge syncopation.

Second, we list the measures regarding mathematical irregularity. IOI histogram measures for entropy (*ioi-g-h*, *ioi-l-h*), standard deviation (*ioi-g-sd*, *ioi-l-sd*), and maximum bin height (*ioi-g-mm*, *ioi-l-mm*) were used to determine the complexity of both global (full) IOIs [24] and local (relative, adjacent) IOIs [18]. Also, pertaining to entropy calculations are the Coded Element Processing System (*ceps*) [26], H(*k*-span) complexity (*hk*) [25], and the H(run-span) complexity (*hrun*) [25], which all measure the uncertainty [5] of obtaining sub-patterns in a rhythm. The directed swap distance (*dswap*, *dswap2*, *dswap4*, *dswap8*) [1] computes the minimum weight of a linear assignment between onsets of a rhythm and a meter with an onset at every second, fourth, or eighth pulse, and also the average over each meter. Two other measures, Rhythmic Oddity (*oddity*) [22] and Off-Beatness (*off-beatness*) [22] take a geometric approach.

Third, those measures which do not easily fall into a category are listed. These include the Lempel-Ziv compression measure (*lz*) [12], Tanguiane's [20] complexity measure, which looks at sub-patterns at each metrical beat level, and Pressing's Cognitive Complexity measure (*pressing*) designed on the basis of music theory principles, which generates rhythmic patterns at each metrical beat, assigning appropriate weights to special patterns [16]. Furthermore, Tanguiane's measure uses the max (*tmmax*) and average (*tmax*) complexities over different metrical beat levels. In addition, derivatives (*tmuavg*, *tmumax*) without the restriction of sub-patterns starting with an onset, were tested.

3 EXPERIMENTAL DATA

The measures of complexity in § 2, were compared using three rhythm data sets. Each data set had been compiled to test human judgements regarding the perceptual, meter, and performance complexities of the rhythms. The first data set shown in Table 1 was synthesized by Povel and Essens in 1985 [15] and then later studied by Shmulevich and Povel in 2000 [17]. The second data set shown in Table 2 was created by Essens in 1995 [2]. The third data set shown in Table 3 was generated by Fitch and Rosenfeld in 2007 [4]. In addition to the rhythms themselves, the results of several human performance complexity measures used in this work are contained in Tables 1, 2, and 3. In the following we describe the methodologies of Povel and Essens [15], Shmulevich and Povel [17], Essens [2], and Fitch and Rosenfeld [4], used to obtain the human judgements of complexity.

3.1 Povel and Essens 1985

Previous work by Povel and Essens [15] studied the reproduction quality of temporal patterns. The rhythms, shown in Table 1, were presented to the participants in random order. For each presentation, the participant was asked to listen to the pattern, and then reproduce the pattern by tapping [15]. Once the participant had felt they could reproduce the rhythm, they stopped the audio presentation and proceeded to then tap the pattern they just heard, repeating the pattern 4 times. Afterwards, they could choose to move to the next rhythm or repeat the one they had just heard [15]. From this experiment, we derive an empirical measure for the reproduction difficulty of temporal patterns; i.e., rhythm performance complexity. This was based on Povel and Essens' *mean deviation percentage* which calculates the amount of adjacent IOI error upon reproduction [15]. See column 3 of Table 1.

3.2 Shmulevich and Povel 2000

Shmulevich and Povel [17] studied the perceptual complexity of rhythms using the same data as Povel and Essens [15]. All participants were musicians with musical experience averaging 9.2 years [17]. A pattern was repeated four times before the next was randomly presented. The resulting perceptual complexity in column 4 of Table 1 represents the average complexity of each rhythm across all participants.

3.3 Essens 1995

A study of rhythm performance complexity was conducted by Essens [2]. The rhythms used for that study are shown in Table 2. The procedure Essens used to test the reproduction accuracy of rhythms was very similar to that of Povel and Essens [15]. We use the mean deviations of Essens to rank the rhythms by increasing complexity, as seen in column 3

of Table 2. Essens also studied the perceptual complexity of rhythms [2]. Participants were asked to judge the complexity of each rhythm in Table 2 on a 1 to 5 scale where 1 means very simple and 5 means very complex [2]. Note that some participants had been musically trained for at least 5 years. The order of the patterns was random. The perceptual complexity in column 4 of Table 2 is the average complexity over the judgements from each subject.

3.4 Fitch and Rosenfeld 2007

Most recently, Fitch and Rosenfeld [4] conducted an experimental study of metric beat-tracking or, in their terminology, pulse-tracking (i.e., rhythmic *meter complexity*) and rhythm reproduction (i.e., *performance complexity*). The rhythms used in the experiments are shown in Table 3. These rhythms were generated in such a way as to vary the amount of syncopation among the rhythms, as measured by the Longuet-Higgins and Lee syncopation measure [14].

The metric beat-tracking experiment yielded two measures of meter complexity [3]. The first pertained to how well participants could tap a steady beat (*beat tapping error adjusted* for tempo) when different rhythms were played. The second counted the number of times (*number of resets*) the participant tapped the metric beat exactly in between the points where the metric beat should be [4]. The values are shown in columns 3 and 4 of Table 3. The second experiment for rhythm reproduction accuracy was interleaved with the metric beat-tracking experiment. Hence the subject now taps the rhythm just heard from experiment 1 while the computer provides the metric beat [4]. The adjacent IOI error of the target and reproduced rhythms gives a performance complexity shown in column 5 of Table 3.

4 RESULTS

We adhered to the following procedure to validate the complexity measures in § 2 using the three rhythm data sets. The complexity scores were obtained using the rhythms as input for each measure. The Spearman rank correlation coefficients [11] between all pairs of rankings of the rhythms according to the computational and empirical measures for each rhythm data set were calculated.

Phylogenetic trees were used to visualize the relationships among the correlation coefficients. This technique has proved to be a powerful analytical tool in the computational music domain [1, 8, 21, 22, 23]. The program SplitsTree [9] was used to generate the phylogenetic trees using the BioNJ algorithm [6]. Figures 1, 2, and 3, picture the phylogenetic trees where the distance matrix values are the correlation coefficients subtracted from one. Each tree yields a fitness value greater than or equal to 94.0 on a 100.0 scale. The least-squares fitness is the ratio of A/B where A is the sum of the squared differences between the geodesic distances

No.	Rhythm	Performance Complexity Povel and Essens	Perceptual Complexity Shmulevich and Povel
1	xxxxx . xx . x . . .	5	1.56
2	xxx . x . xxx	1	2.12
3	x . xxx . xxx	0	2.08
4	x . x . xxxxxx	2	1.88
5	x . . xx . x . xxxxxx . . .	3	1.80
6	xxx . xxx . xx	9	2.44
7	x . xxxx . xx	7	2.20
8	xx . . xxxxxx . x	4	2.56
9	xx . . x . xxx . xxx	14	3.00
10	x . xxx . xxx	18	2.04
11	xxx . xx . . xx	19	2.76
12	xx . xxx	15	2.72
13	xx . xx . xxx	13	3.00
14	xx . . xx . xx . xxx	27	3.16
15	x . . xxx . xxx . xx	10	2.04
16	xx . xxx . xx	11	2.88
17	xx . xxx . xxx	17	2.60
18	xx . xxx . . xx	22	2.60
19	xx . . xx . xxx	21	2.64
20	xx . . . xx . xxx . xx . . .	25	3.24
21	xxxxx . xx . x	29	3.08
22	xxxx . x . . xxx	20	3.04
23	xxx . . xx . xxx	16	3.04
24	x . . xxx	6	2.56
25	x . . . xxx . xxx	8	2.56
26	xxxx . x	26	2.84
27	xx . xxx	23	3.60
28	xx xxx . xxx	32	2.68
29	x . xxx	28	3.28
30	x . . xxx	21	3.08
31	xxxx . xxx	30	3.52
32	xxxx . . xx . xx	31	3.60
33	xx . xxx	24	3.04
34	xx . x . . xxx	33	2.88
35	x . x . . xxx . xxx	12	3.08

Table 1. Thirty-five rhythms from Povel and Essens with the Performance Complexity and Perceptual Complexity.

No.	Rhythm	Performance Complexity Essens	Perceptual Complexity Essens
1	xxx . xxx . xxx . xx . . .	0	2.2
2	xxx . x . xxx . xxx	8	3.1
3	x . xxx . xxx	4	3.2
4	x . . xxx . . xx . xxx	19	2.9
5	xxx . xxx . xx . xx	2	2.2
6	xxx . x . . xx . x . xxx . .	7	3.1
7	xxxxxxxx . xxx . xxx	10	2.6
8	xxx . xxx	5	4.2
9	xxxxxxxx . xx . xxx	13	2.9
10	x . . x . x . xxx . xx	6	2.8
11	xxxxxxxx . xxx . x . x . . .	1	3.1
12	xxx . xx . . . x . x	3	2.5
13	x . . xxx . . xx	20	3.5
14	x . xxx . xxx . xxx	12	2.5
15	x . . xxx . xxx . xxx	14	2.4
16	x . . xxx . xxx . xx	11	3.0
17	xx . xxx . xxx . x	17	3.0
18	x . . xxx . xxx	18	3.1
19	x . . x . xx . xxx . xxx . . .	22	2.4
20	xx . xxx . xx . xx	16	3.2
21	xx . xxx . xxx . xxx	15	2.4
22	xx . . xx . xxx . xxx	11	2.9
23	x . . xx . xxx . xxx	21	2.7
24	xx . xxx	9	3.8

Table 2. Twenty-four rhythms from Essens with Performance Complexity and Perceptual Complexity.

between pairs of leaves in the tree, and their corresponding distances in the distance matrix, and B is the sum of the squared distances in the distance matrix. Then this value is subtracted from 1 and multiplied by 100 [27]. Note that the phylogenetic tree is used here as a visualization tool, and not in order to obtain a phylogeny of complexity measures.

5 DISCUSSION AND CONCLUSION

There are several noteworthy observations common to all three data sets. The syncopation measures form a tight cluster far from the other clusters. The human performance mea-

Povel and Essens - Shmulevich and Povel - Spearman Rank Correlation Coefficients - BioNJ LSFit 97.6

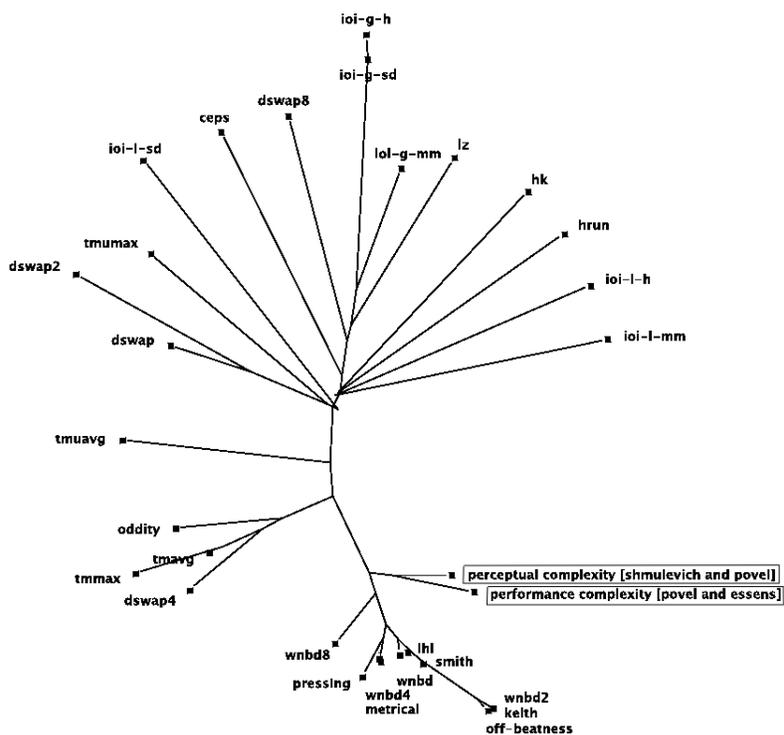


Figure 1. BioNJ of measures compared to Shmulevich and Povel and Povel and Essens human judgements.

Essens - Spearman Rank Correlation Coefficients - BioNJ LSFit 94.0

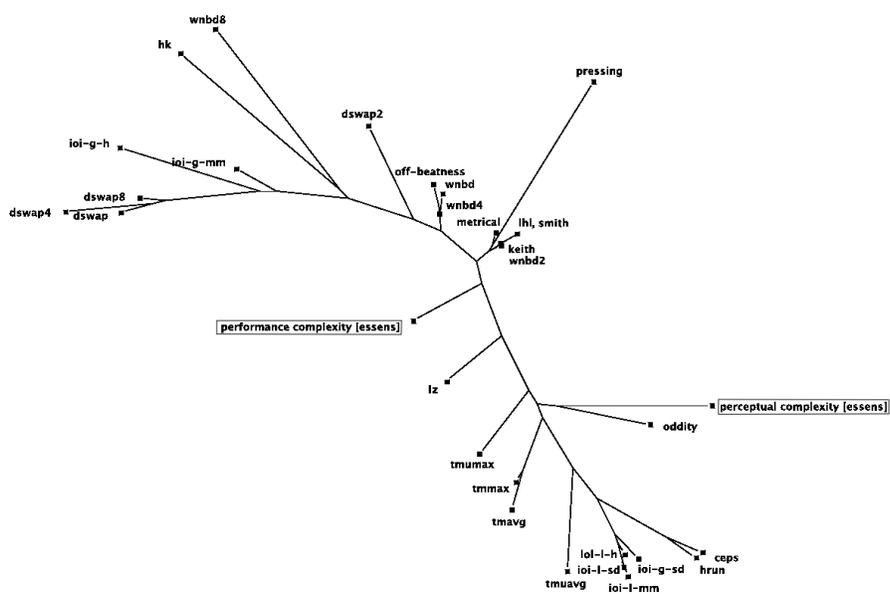


Figure 2. BioNJ of measures compared to Essens' human judgements.

Fitch and Rosenfeld - Spearman Rank Correlation Coefficients - BioNJ LSFit 97.4

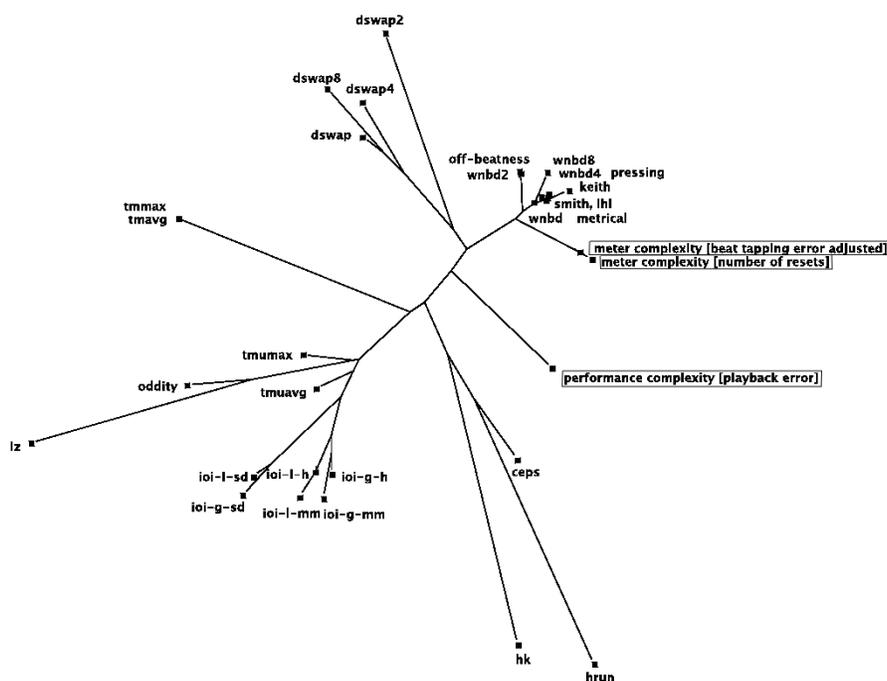


Figure 3. BioNJ of measures compared to Fitch and Rosenfeld’s human judgements.

No.	Rhythm	Meter Complexity Beat Tapping Adjusted	Meter Complexity Number of Resets	Performance Complexity Play-back Error
1	x x x	0.075	2.500	0.138
2	x x x	0.082	2.250	0.145
3	x x x	0.075	2.313	0.153
4	x x x	0.119	8.750	0.257
5 x x	0.103	5.500	0.133
6 x x	0.082	3.063	0.235
7	x x x	0.112	6.000	0.215
8	x x x	0.110	5.188	0.208
9	x x x	0.141	6.938	0.250
10	x x x	0.144	10.375	0.171
11 x x	0.130	6.875	0.220
12 x x	0.124	6.438	0.226
13 x x	0.130	6.965	0.387
14 x x	0.159	11.688	0.239
15 x x	0.172	13.688	0.485
16	x x x	0.085	2.625	0.173
17	x x x	0.077	2.313	0.179
18 x x	0.077	2.438	0.182
19	x x x	0.074	1.938	0.252
20 x x	0.098	3.375	0.142
21	x x x	0.161	11.063	0.305
22 x x	0.129	8.500	0.321
23	x x x	0.145	7.375	0.320
24 x x	0.134	7.188	0.265
25 x x	0.146	8.625	0.176
26 x x	0.118	6.500	0.326
27 x x	0.117	6.188	0.368
28 x x	0.154	10.813	0.344
29 x x	0.191	15.750	0.185
30 x x	0.164	11.938	0.158

Table 3. Thirty rhythms from Fitch and Rosenfeld with Meter Complexity and Performance Complexity.

asures fall in the same cluster as the syncopation measures. The complexity measures based on statistical properties of the inter-onset-interval histograms appear to be poor predictors of syncopation or of human performance complexity.

There are also some important differences between the three figures. The overall appearance of clusters is much stronger in Figure 3 than in the other two. This is perhaps due to the fact that the rhythms used in Figure 3 are much more realistic and sparser than the rhythms used in Figures 1 and 2. Similarly, the six IOI (inter-onset-interval) measures are scattered in Figures 1 and 2, but are in one cluster in Figure 3. The cognitive complexity measure of Pressing, designed on the basis of principles of music perception falls squarely in the group of syncopation measures in Figures 1 and 3. However, in Figure 2, although it falls into the syncopation cluster, it is quite distant from the other measures, probably because of the great density of the rhythms in this data set. Also worthy of note is a comparison of the human meter complexity measures with the human performance (playback) measure. In Figure 3 we see that the meter complexity is considerably closer to the syncopation measures than the play-back performance measure. This suggests that the mathematical syncopation measures are better predictors of human meter complexity than performance complexity.

6 ACKNOWLEDGEMENT

The authors would like to thank W. T. Fitch for making their data set available.

7 REFERENCES

- [1] J. M. Díaz-Báñez, G. Farigu, F. Gómez, D. Rappaport, and G. T. Toussaint. El compás flamenco: a phylogenetic analysis. In *BRIDGES: Mathematical Connections in Art, Music and Science*, Jul 2004.
- [2] P. Essens. Structuring temporal sequences: Comparison of models and factors of complexity. *Perception and Psychophysics*, 57(4):519–532, 1995.
- [3] W. T. Fitch. Personal communication, 2007.
- [4] W. T. Fitch and A. J. Rosenfeld. Perception and production of syncopated rhythms. *Music Perception*, 25(1):43–58, 2007.
- [5] W. R. Garner. *Uncertainty and Structure as Psychological Concepts*. John Wiley & Sons, Inc., 1962.
- [6] O. Gascuel. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution*, 14(7):685–695, 1997.
- [7] F. Gómez, A. Melvin, D. Rappaport, and G. T. Toussaint. Mathematical measures of syncopation. In *BRIDGES: Mathematical Connections in Art, Music and Science*, pages 73–84, Jul 2005.
- [8] F. Gómez, E. Thul, and G. T. Toussaint. An experimental comparison of formal measures of rhythmic syncopation. In *Proceedings of the International Computer Music Conference*, pages 101–104, Aug 2007.
- [9] D. H. Huson and D. Bryant. Applications of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2):254–267, 2006.
- [10] M. Keith. *From Polychords to Pólya: Adventures in Musical Combinatorics*. Vinculum Press, 1991.
- [11] M. Kendall and J. D. Gibbons. *Rank Correlation Methods, Fifth Edition*. Oxford Univ. Press, New York, 1990.
- [12] A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Transactions on Information Theory*, IT-22(1):75–81, 1976.
- [13] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, 1983.
- [14] H. C. Longuet-Higgins and C. S. Lee. The rhythmic interpretation of monophonic music. *Music Perception*, 1(4):424–441, 1984.
- [15] D.-J. Povel and P. Essens. Perception of temporal patterns. *Music Perception*, 2:411–440, 1985.
- [16] J. Pressing. Cognitive complexity and the structure of musical patterns. <http://www.psych.unimelb.edu.au/staff/jp/cog-music.pdf>, 1999.
- [17] I. Shmulevich and D.-J. Povel. Measures of temporal pattern complexity. *Journal of New Music Research*, 29(1):61–69, 2000.
- [18] I. Shmulevich, O. Yli-Harja, E. Coyle, D.-J. Povel, and K. Lemström. Perceptual issues in music pattern recognition: complexity of rhythm and key finding. *Computers and the Humanities*, 35:23–35, February 2001.
- [19] L. M. Smith and H. Honing. Evaluating and extending computational models of rhythmic syncopation in music. In *Proceedings of the International Computer Music Conference*, pages 688–691, 2006.
- [20] A. S. Tanguiane. *Artificial Perception and Music Recognition*. Springer-Verlag, 1993.
- [21] G. T. Toussaint. A mathematical analysis of African, Brazilian, and Cuban clave rhythms. In *BRIDGES: Mathematical Connections in Art, Music and Science*, pages 157–168, Jul 2002.
- [22] G. T. Toussaint. Classification of phylogenetic analysis of African ternary rhythm timelines. In *BRIDGES: Mathematical Connections in Art, Music and Science*, pages 23–27, Jul 2003.
- [23] G. T. Toussaint. A comparison of rhythmic similarity measures. In *Proc. International Conf. on Music Information Retrieval*, pages 242–245, Universitat Pompeu Fabra, Barcelona, Spain, October 10-14 2004.
- [24] G. T. Toussaint. The geometry of musical rhythm. In Jin Akiyama, Mikio Kano, and Xuehou Tan, editors, *Proc. Japan Conf. on Discrete and Computational Geometry*, volume 3742 of *Lecture Notes in Computer Science*, pages 198–212. Springer Berlin/Heidelberg, 2005.
- [25] P. C. Vitz. Information, run structure and binary pattern complexity. *Perception and Psychophysics*, 3(4A):275–280, 1968.
- [26] P. C. Vitz and T. C. Todd. A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, 75(6):433–449, Sep 1969.
- [27] R. Winkworth, D. Bryant, P. J. Lockhart, D. Havell, and V. Moulton. Biogeographic interpretation of splits graphs: least squares optimization of branch lengths. *Systematic Biology*, 54(1):56–65, 2005.