The Pairwise Variability Index as a Measure of Rhythm Complexity[1]

Godfried T. Toussaint

I. INTRODUCTION

The normalized pairwise variability index (nPVI) is a measure of the average variation (contrast) of a set of distances (durations) that are obtained from successive adjacent ordered pairs of events. It was originally conceived for measuring the rhythmic differences between languages on the basis of vowel length (Grabe and Low 2002), and several successful applications in this domain have been realized (Gibbon and Gut 2001; Olivo 2011). Asu and Nolan (2005) concluded that the nPVI of the syllable durations was appropriate for measuring the complexity of the Estonian language. It has also been applied to the determination of the cognitive complexity of using text-entry systems (Sandness and Jian 2004). A review of the history, rationale, and application of the nPVI to the study of languages is given by Nolan and Asu (2009). More recently, the measure has also been employed successfully to compare speech rhythm with rhythm in music (McGowan and Levitt 2011; London and Jones 2011; Patel and Daniele 2003; Huron and Ollen 2003; Daniele and Patel 2004). It has been suggested by London and Jones (2011) that the nPVI could become a useful tool for musical rhythm analysis as such. Indeed, it has already informed several questions about music as well as the interplay between music and language (McDonough et al. 2007). It has been used successfully to compare rhythm in musical scores with their

[1] This is an extended version of a paper that appeared in the *Proceedings of the 12th International Conference on Music Perception and Cognition* (ICMPC), and *8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (ESCOM), in Thessaloniki, Greece, on July 23-28, 2012, pp. 1001–1008, under the title: “The pairwise variability index as a tool in musical rhythm analysis.”

performances (Raju, Asu and Ross 2010), and it can distinguish between compositional styles in nineteenth-century French and German art song (VanHandel 2006; VanHandel and Song 2009; see also Dalla Bella and Peretz 2005). Furthermore, using English and Japanese as a case study, it has been demonstrated that a composer's native language may influence her or his musical compositions (Sadakaka et al. 2004). In the domain of psychology, the nPVI has provided evidence that there exists a "common code" of emotional expression in speech and music (Quinto 2013).

One goal of the present study is to determine how well the nPVI correlates with various dimensions of musical rhythm complexity, ranging from human performance and perceptual complexities to mathematical measures of metric complexity and rhythm irregularity. A second goal is to determine to what extent the nPVI is capable of discriminating between short, symbolically notated, durational patterns that occur in musical rhythms, metric frames, genres, styles, and cultures, as well as across strictly non-musical "rhythms" such as the highly irregular mark patterns of Golomb rulers (discussed in Section III). It is shown that the nPVI suffers from several shortcomings when it comes to modelling metric and rhythmic complexity in the context of short symbolic rhythmic patterns such as sub-Saharan African bell patterns, Arabic rhythms, Rumanian dance rhythms, and Indian *talas*. Nevertheless, comparisons with previous experimental results reveal that the nPVI correlates moderately with human performance complexity. The index is also able to discriminate between almost all the families of rhythms tested. However, no highly significant differences were found between the nPVI values for binary and ternary musical rhythms, partly mirroring the findings by Patel and Daniele (2003) for language rhythms. In addition, a modification of the nPVI is proposed that incorporates the underlying meter, and which correlates highly with two measures of human performance complexity for musical rhythms that are syncopated.

II. CHANGE AS A MEASURE OF COMPLEXITY

Nick Chater (1999, 287) suggests that judgments of complexity are akin to judgments of irregularity. Since a rhythm consists of a pattern of inter-onset intervals, one way to measure the irregularity of the rhythm is by measuring the irregularity of the intervals that make up the rhythm. The *standard deviation*, used in statistics and probability, is a widely used measure of the dispersion of a random variable. When the random variable is the size of the inter-onset intervals it becomes a measure of irregularity, and has thus been used frequently in speech and language studies. However, musical rhythm and speech are not static; they are processes that unfold in time, and the standard deviation measure "lifts" the intervals out of their original order, disregarding the relationships that exist between adjacent intervals. A better measure of variability across time should incorporate change in the durations of adjacent intervals. Measures of change have been used to characterize the complexity of binary sequences in the visual perceptual domain. Indeed, Psotka designed a measure called *syntely* to gauge how much the structure of the early portions of a sequence influence the terminal sections, or "the strength of stimulus continuation" (1975, 436). Furthermore, Aksentijevic and Gibson (2003, 2012) characterized psychological complexity as change: "Structural information is contained in the transition from one symbol (or element) to another and not in the symbols themselves" (2012, 1). The nPVI is a measure of variability that attempts to capture this notion of change. Grabe and Low (2002) define the nPVI for a rhythm with adjacent inter-onset intervals (IOIs) as

$$nPVI = \left(\frac{100}{m-1}\right)\sum_{k=1}^{m-1}\left|\frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2}\right|$$

where *m* is the number of adjacent vocalic intervals in an utterance, and d_k is the duration of the k^{th} interval. Translating this terminology to the musical rhythmic domain converts the vocalic intervals to the adjacent IOIs. Notably, Patel (2008, 133) laments the use of the term “variability” for the nPVI because, as he rightly points out, it is fundamentally a measure of temporal *contrast* between adjacent durations, rather than variation, and the two measures are not monotonically related (i.e., an increase (or decrease) in one does not imply an increase (or decrease) in the other). Consider the two durational patterns A = [2-5-1-3] and B = [1-3-6-2-5]. The rhythm A has a lower standard deviation than B (1.708 versus 2.074), whereas it has a higher nPVI value (106.3 versus 88.1). In spite of anomalies such as these, the results presented below indicate that in many musical contexts the two measures are highly and statistically significantly correlated.

To clarify the close relationship that exists between the standard deviation and the nPVI, consider a sequence of three adjacent inter-onset intervals *a*, *b*, *c* such that $a + b + c = 1$, and $a = c$, and refer to Figures 1 and 2. The two measures are plotted as a function of *b*, the duration of the middle interval. Both measures take on their minimum value (zero) when the three intervals are equal, and both take on their maximum values when *b* is either zero or one. Also both functions decrease monotonically from $b = 0$ to 1/3, and increase monotonically from $b = 1/3$ to 1. However, whereas the standard deviation varies linearly, the nPVI does not. The nPVI function is slightly convex from $b = 0$ to 1/3, and slightly concave from $b = 1/3$ to 1. The only substantial difference between the two curves is in the regions near their maximum values at the extremes of *b*. The nPVI takes on the same value (200) for $b =$

0 and $b = 1$, whereas the standard deviation is almost twice as large for $b = 1$ than for $b = 0$. Consider two IOI sequences: A = (0.15, 0.7, 0.15) and B = (0.45, 0.1, 0.45). The nPVI regards these two sequences as having almost equal variability: nPVI(A) = 130 and nPVI(B) = 127. However, their standard deviations are quite different: SD(A) = 0.32 and SD(B) = 0.20.

Figure 1. The standard deviation as a function of b for $a+b+c=1$ and $a=c$

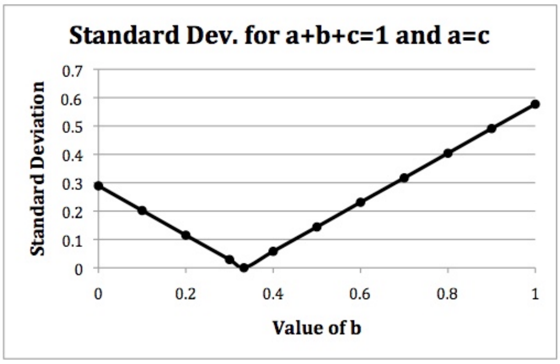
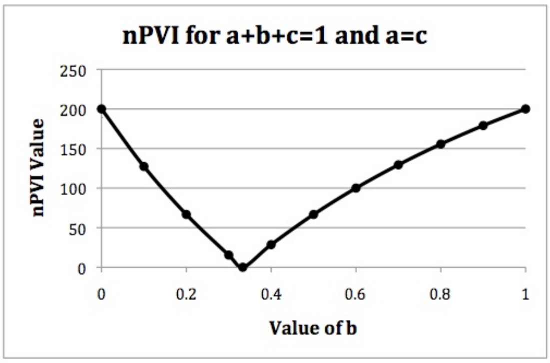


Figure 2. The nPVI as a function of b for $a+b+c=1$ and $a=c$



III. METHOD AND DATA

Several sets of rhythms (both musical and non-musical), notated as binary sequences, were collected, including various families of synthetic rhythms (random and systematic), rhythms from India, the sub-Saharan African diaspora, the Arab world, Rumania, as well as a collection of Golomb ruler patterns. Some had been previously evaluated experimentally according to several measures of human performance and perceptual complexities. The nPVI values and standard deviations of the IOIs of all rhythms were calculated. These values yielded, for each data set, two orderings (rankings) of the rhythms. These orders were then compared with those produced by the remaining measures using Spearman rank correlation coefficients and Kolmogorov-Smirnov tests.[2]

The Data Sets Used

Several researchers have carried out listening experiments with collections of artificially generated rhythms in order to test several hypotheses about the mental representations of rhythms. The experiments done by Povel and Essens (1985), Shmulevich and Povel (2000), Essens (1995), and Fitch and Rosenfeld (2007) used three data sets of rhythms that provided measures of human perceptual and performance complexities for each rhythm. The three data sets are briefly described in the following. For further details and listings of all the rhythms (in box notation) the reader is referred to the original papers. All the rhythms in these three data sets consist of time spans (measures, cycles) of sixteen unit pulses.

The Povel-Essens data. This data comprises thirty-five rhythms, all of which contain nine attacks (onsets). They all start with an attack on the first pulse, and end with a long interval [x . . .]. The rhythms are made up of all possible permutations of the nine IOIs {1,1,1,1,1,2,2,3,4}, and do not resemble the rhythmic timelines used in

[2] The Spearman rank correlation coefficient is a measure of how well two ranked orders of elements agree with each other; that is, the degree to which two variables are related monotonically (Spearman 1904). The Kolmogorov-Smirnov test is used here to measure the distance between two empirically estimated distributions (Conover 1971, 295–301).

traditional music. Every rhythm in the collection has five intervals of duration 1, two of duration 2, one of duration 3, and one of duration 4.

The Essens data. This data set consists of twenty-four rhythms, in which the number of onsets varies between eight and thirteen, and is generally larger than that of the Povel-Essens rhythms. All the rhythms also start with an attack on their first pulse. Like the rhythms in the Povel-Essens data, these rhythms bear little resemblance to the rhythms used as timelines in musical practice.

The Fitch-Rosenfeld data. This data set consists of thirty rhythms, in which the number of onsets is smaller than in the rhythms of the preceding two data sets; six rhythms have four onsets and the rest have five. Also noteworthy is that unlike the other two data sets, seventeen rhythms possess a silent downbeat (i.e., they start on a silent pulse; Huron 2006, 200). Furthermore, in contrast to the two data sets described above, these rhythms were "composed manually" in such a way as to vary the amount of syncopation present in the rhythms, as measured by Fitch and Rosenfeld's implementation of the syncopation measure of Longuet-Higgins and Lee (1984). See also the related relevant work by Smith and Honing (2006). Fitch and Rosenfeld (2007) did not indicate, and perhaps did not appear to realize, that most of the rhythms generated (or their cyclic rotations) are in fact rhythmic patterns used in practice in sub-Saharan African and Indian music. Therefore this data set, although generated algorithmically, differs widely from the Povel-Essens and Essens data sets in that it reflects real living rhythms and contains rhythms with a silent first pulse.

Random rhythms. To gain insight into the properties of musical rhythms used in practice it is helpful to compare them with randomly generated rhythms. For part of his study of rhythmic complexity measures, Thul (2008, 58) generated the fifty random 16-pulse rhythms listed in his Table 4.7. The rhythms were obtained by programming a random number generator to simulate "flipping an unbiased coin" sixteen times. Then "heads" was associated with an onset, and "tails" with a silent pulse. The twenty-seven rhythms from this list that started with an onset were chosen for comparison with the other data sets in this study.

North Indian talas. In Indian classical music a *tala* (also *taal* or *tal*) is a cyclically recurring clap pattern of fixed length that corresponds somewhat to the concept of meter in Western music, *timeline* in Sub-Saharan music, or *compás* in the flamenco music of southern Spain. Although talas may not be heard or even perceived, they function as metric frames, were evolutionarily selected by music scholars, and thus form part of the Indian music theory consciousness. It is therefore interesting, especially from a scientific point of view, to compare them with similar concepts in other cultures. Clayton (2000) provides an in-depth analysis of talas and their role in North Indian classical music. The twelve North Indian talas used here were taken from his Example 5.1 (58–59). For a detailed comparison of North Indian talas and sub-Saharan African timelines, see Thul and Toussaint (2008b).

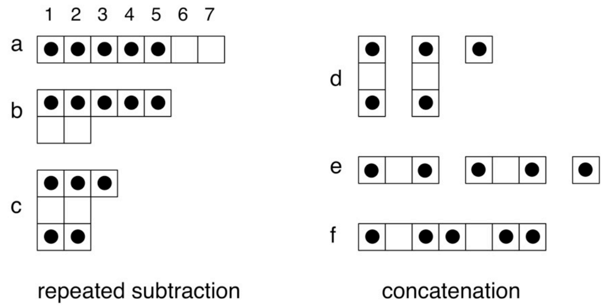
South Indian talas. The Carnatic music of South India is also based on rhythmic cycles. One of these systems consists of thirty-five *sulaadi talas* (Morris 1998). Five of these talas are made up of single durations; therefore, the nPVI and standard deviation are not defined for them. The data used in this study consisted of the other thirty talas.

Desi talas. A thirteenth-century Indian manuscript written by Sarngadeva lists 130 talas called *desi talas*, which may be found in Johnson (1975, 194; see also Morris 1998). Fourteen of these are isochronous rhythms that yield nPVI and standard deviation values of zero. Since these entries create ties in the rankings and inflate the Spearman rank correlation coefficients, they were removed, leaving 116 irregular desi talas for the comparisons listed in Table 1. The desi talas have the greatest range of pulses ($3 \leq n \leq 71$) and onsets ($2 \leq k \leq 19$) of all the data sets used in this study. Thus they provide a good data set with which to study how the nPVI varies as a function of *k* and *n*. Not surprisingly, higher values of *n* tend to have more onsets. The Spearman rank correlation (denoted by *r* throughout) bears this out: $r = 0.68$ with $p < 0.000001$, where *p* denotes the significance level. The number of onsets relative to the number of pulses, also called the note density (Cerulo 1988), may be considered to be a mild contributing factor of rhythm complexity. However, for the desi talas the nPVI and *k* yield a low but significant negative correlation: $r = -0.2$ with $p < 0.015$. It appears that an increase in the number of onsets tends to decrease the contrast between the adjacent IOIs. On the other hand, the nPVI does not appear to be correlated with the number of pulses *n* ($r = 0.08$, $p < 0.18$).

Sub-Saharan African timelines. Agawu (2006, 1) defines a timeline as a "bell pattern, bell rhythm, guideline, time keeper, topos, and phrasing referent," and characterizes it as a "rhythmic figure of modest duration that is played as an ostinato throughout a given dance composition." Most timelines used in the sub-Saharan African diaspora employ a cycle of twelve (ternary) or sixteen (binary) pulses. Perhaps the most distinctive of these are the binary and ternary "signature" timelines with five onsets and durational patterns [3-3-4-2-4] and [2-2-3-2-3], respectively (Toussaint 2011). In the first study thirty-nine ternary and thirty-seven binary sub-Saharan African timelines were taken from the papers by Rahn (1987, 1996) and Toussaint (2005), as well as the articles referenced therein. The ternary rhythms had twelve pulses and the number of onsets varied between four and nine. The binary rhythms had sixteen pulses and the onsets varied between three and ten. In the second study only the timelines that had five onsets were used.

Euclidean rhythms. One of the oldest and most well known algorithms in the field of computer science, identified in Euclid's *Elements* (circa 300 BCE) as Proposition II in Book VII, is known today as the Euclidean algorithm. It was designed to compute the greatest common divisor of two given integers (see Franklin 1956). Donald Knuth calls this algorithm the "granddaddy of all algorithms, because it is the oldest nontrivial algorithm that has survived to the present day" (1998, 335). The idea behind the Euclidean algorithm is remarkably simple: repeatedly replace the larger of the two numbers k and n by their difference, until both are equal. The final number thus obtained is the greatest common divisor of k and n. Consider the two numbers $k = 5$ and $n = 8$. Subtracting 5 from 8 yields 3; 5 minus 3 equals 2; 3 minus 2 equals 1; and finally, 2 minus 1 equals 1. Therefore, the greatest common divisor of 5 and 8 is 1. In Toussaint (2005) it was shown that by associating n with the number of pulses in the cycle, and k with the number of onsets (attacks), the Euclidean algorithm may be used to generate most rhythm timelines used in traditional music found all over the world. The key lies in extracting not the answer to the original problem, but rather the structure observed in the repeated subtraction process used to obtain the answer. This process is illustrated in Figure 3 with $k = 5$ and $n = 7$. The figure visually illustrates the following operations performed repeatedly by the Euclidean algorithm. First the onsets and silent pulses are ordered one after the other as in (a). Then in the repeated subtraction phase, the silent pulses are moved to the positions as shown in (b). If there are more silent pulses than onsets then the number of silent pulses moved is the same as the number of onsets. This subtraction process is continued until the number of columns remaining is either zero or one, as in (c). Note that if k divides evenly into n the number of columns remaining will equal zero and the resulting rhythm will be isochronous (i.e., not just maximally even, but

Figure 3. Generating a Euclidean rhythm with $k = 5$ and $n = 7$



perfectly even). The resulting columns are then concatenated as in (d) and (e) to obtain the final rhythm in (f). In other words, the pulses in (c) are read from top to bottom and left to right in order to spread the onsets out as evenly as possible. This particular Euclidean rhythm denoted by E(5,7) = [x . x x . x x] = (21211) is the *Nawakhat* pattern, a popular Arabic rhythm (Standifer 1988). In Nubia it is called the *Al Noht* rhythm (Hagoel 2003). Note that any rotation of this rhythm such as (12112) or (21121), and so forth, is also a Euclidean rhythm. We say that the algorithm in fact generates a Euclidean necklace. In theoretical computer science, Euclidean rhythms, also known as Euclidean strings (Ellis et al. 2003), have been discovered independently in different contexts, such as calendar leap year calculations, drawing digital straight lines, and word theory. In music theory, they are called *maximally even* sets.[3] Here the terms maximally even and Euclidean are used interchangeably. In the experiments described here the forty-six Euclidean rhythms used were taken from Toussaint (2005) and the references therein. Since Euclidean rhythms may be generated by such a simple rule, it follows that they have very low Kolmogorov (or information theoretic) complexity (see Chaitin 1974, and Lempel and Ziv 1976).

[3] For a sampling of this literature, see Amiot (2007), Clough and Douthett (1991), Douthett and Krantz (2007) and Toussaint (2005, 2013).

Furthermore, by their property of maximally evenly distributed onsets, the rhythms tend to maximize repetitiveness and minimize contrast. They are, therefore, expected to have relatively low nPVI values, and furnish an extreme data set for comparison with the other data sets. This property also permits the robustness of the nPVI to be tested across widely differing data sets.

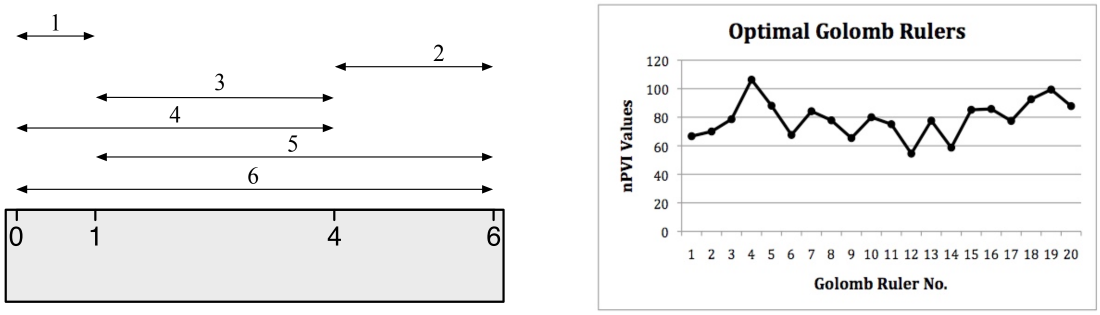
Rumanian folk dance rhythms. Proca-Ciortea (1969) investigated over 1,100 Rumanian folk dances from which fifty-six rhythms were extracted. The rhythms, which are mostly binary, consist mainly of eighth and quarter notes. Six of these rhythms were either isochronous or had a silent first pulse, and were deleted. The remaining fifty rhythms were used in this study.

Arabian wazn. Rhythmic patterns in Arabian music are known as *wazn*. They may be compared to sub-Saharan African timelines in structure and function, although their IOI patterns are quite different. Whereas the African timelines use time spans (measures) that are composed predominantly of twelve and sixteen pulses, the Arabian wazn employ a wide variety of different values. The data used here were composed of nineteen wazn taken from the book by Touma (1996). The longest wazn was the *samah* consisting of nineteen onsets in a time span of thirty-six pulses, with an IOI pattern given by [2-1-1-4-1-1-1-1-2-4-2-1-1-1-1]. Compared to the African timelines, the Arabian wazn appear to be more irregular.

Optimal Golomb rulers. A Golomb ruler, also referred to as a *Sidon set* by Erdős and Turán (1941), is a ruler that has “few” marks, and which thus permits measuring distances only between pairs of these marks. Furthermore, to increase the versatility of a Golomb ruler with a fixed number of marks, it is desirable to be able to measure as many distinct distances as possible.[4] The problem arises in several

[4] See Alperin and Drobot (2011) for a clear and accessible introduction to the theory of Golomb rulers.

Figure 4. A perfect optimal Golomb ruler with four marks (left) and the nPVI for the 20 shortest optimal Golomb rulers (right)



applications such as the need for distributing expensive radio telescope elements across a stretch of land so as to better receive signals from outer space. An *optimal* Golomb ruler with *k* marks is one such that no other shorter Golomb ruler with *k* marks exists. Furthermore, if the Golomb ruler measures all the distances ranging from one to the length of the ruler (one time each) it is called *perfect*. For example, the ruler shown in Figure 4 (left) with marks at points 0, 1, 4, and 6 is a perfect and optimal Golomb ruler. The length of the ruler is six, the pairwise distances realized are the integers in the set {1, 2, 3, 4, 5, 6}, and it yields the "rhythm" [x x . . x .] with durational IOI pattern [1-3-2]. The twenty shortest optimal Golomb rulers, starting with the 3-mark ruler (0, 1, 3), were obtained from Shearer (2012). The motivation for using these rulers in the experiments is that, as a byproduct of their design to have all their pairwise distances distinct, they yield "rhythms" that are extremely irregular. Since the existence of patterns that emerge from regularities (or repetitions) is considered a necessary ingredient of "beautiful" music (Kivy 1993), Golomb rulers have been used to compose "ugly" music (Rickard 2011).[5] Therefore, Golomb rulers provide another extreme data set useful for the comparative analysis of the nPVI and

[5] See Freeman (1997) for a discussion of the close relationships that exist between Golomb rulers and sequence irregularity.

for gaging the irregularity of rhythms used in different styles of music. The twenty rulers used here had IOIs that varied in number between three and twelve. The ruler with the highest nPVI value of 106.3 in Figure 4 (right) was the fourth in the list, with marks at (0, 2, 7, 8, 11) yielding IOIs [2-5-1-3]. Golomb rulers are closely related to Z-related sets investigated for some time in musical pitch-class theory (Goyette 2012).[6]

Mathematical measures of complexity. In the study of speech rhythm, a measure of variability that has often been used for the vocalic or consonantal interval durations is the classical statistical measure of standard deviation.[7] Since the nPVI was originally proposed as a method for avoiding the drawbacks of the standard deviation, the two measures were compared to determine the extent of their differences. The nPVI was also compared with Keith's measure of metrical complexity. In the musical domain, Michael Keith (1991) proposed a measure of meter complexity based on a hierarchical partition of a meter into sub-meters, and on the frequency of alternations between binary and ternary units within different levels of this hierarchy. In his book, he lists seventy-six metrical patterns with the number of pulses as high as twenty along with their complexity values (1991: 129).

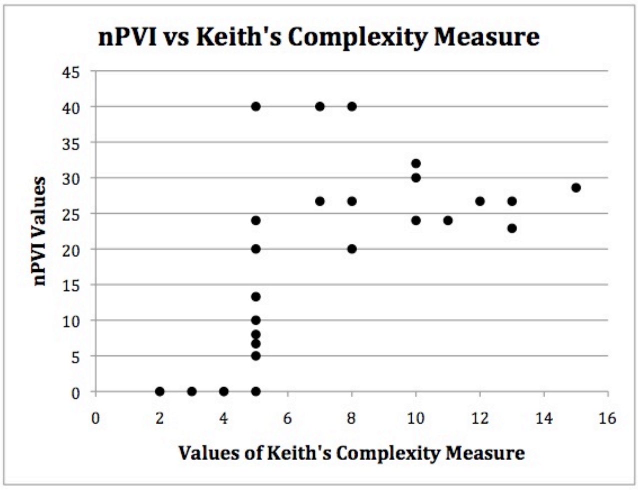
For a string of symbols *S*, Keith's measure of meter complexity, denoted by *C*(*S*), is defined for meters consisting of durational patterns made up of strings of 2s and 3s, such as the African signature pattern also called *fume-fume* [2-2-3-2-3] and the *guajira* flamenco *compás* [3-3-2-2-2] (also Bernstein's "America"; see London 1995). Keith first partitions *S* into any string of disjoint *subunits*. At the lowest level these two rhythms are partitioned into the sub-units [2][2][3][2][3] and [3][3][2][2][2],

[6] Soderberg (2013) recounts personal correspondence with David Lewin in 1994, in which Lewin makes the connection between Golomb rulers and what he called "uniformly flat" Z-sets.
[7] See, e.g., the papers by Ramus, Nespor, and Mehler (1999), and Yoon (2010).

respectively. At this level, the complexity of an individual [2]-unit is 2, and that of a [3]-unit is 3. For a given partition the complexity of the string $C(S)$ is the sum of the complexities of the sub-units, so at this level the complexities of the fume-fume and guajira are the same: 2+2+3+2+3 = 3+3+2+2+2 = 12. Thus at the level of individual units Keith's measure makes no distinction between distributions of IOIs within a cycle. Keith defines a *unit* as one or more *identical contiguous* sub-units. Thus another possible pair of partitions for these two rhythms consists of [2-2][3][2][3] and [3-3][2-2-2], respectively. The complexity value of a unit U consisting of a number of identical sub-units H is defined as $C(U) = max\{\#\text{sub-units}, C(\text{H})\}$, where #sub-units denotes the number of sub-units. For example, the complexity of the unit [2-2] is $max\{2, 2\} = 2$, and that of the unit [2-2-2] is $max\{3, 2\} = 3$. If we denote a partition of S by S_U, then the complexity of a given partition of S into units, denoted by $C(S_U)$, is the sum of the complexities of the units $C(U)$. Therefore, for this partition $C(S_U)$(*fume-fume*) = 2+3+2+3= 10 and $C(S_U)$(*guajira*) = 3+3 = 6. Finally, the complexity of the sequence $C(S)$ is the minimum complexity, minimized over all possible partitions; $C(S) = min\{C(S_U)\}$. In our example the guajira admits several other partitions such as [3-3][2-2][2], [3-3][2][2-2], [3][3][2-2][2], and [3][3][2][2-2], the complexities of which are, respectively, 7, 7, 10, and 10. Therefore, the final complexities are C(*fume-fume*) = 10 and C(*guajira*) = 6.

A scatter-plot of the nPVI values as a function of Keith's complexity measure for the seventy-six metric patterns taken from Michael Keith's book is shown in Figure 5. Although the two measures are highly and significantly correlated ($r = 0.65$, $p < 0.01$), the relationship between the two measures is dominated by vertical and horizontal alignments. Most noticeable is that for a complexity value of 5, there are numerous rhythms with nPVI values ranging from zero to 40.[8]

Figure 5. The nPVI as a function of Keith's measure of metric complexity



IV. RESULTS

Correlation Between nPVI and Other Complexity Measures

The Spearman rank correlation coefficients and statistical significance values obtained by comparing the various complexity measures with the nPVI for all the data sets are listed in Table 1. Those with statistically significant *p*-values are highlighted in bold black, and those that are either uncorrelated or insignificantly correlated are shown in red.

A comparison of the nPVI with the human performance and perceptual complexities obtained with the three data sets of Povel-Essens, Essens, and Fitch-Rosenfeld yields rather curious and mixed results. In these studies, the performance complexity refers to the difficulty of accurately reproducing the rhythm by tapping.

[8] For a detailed comparison of Keith's measure of complexity with scores of other complexity measures, see Thul and Toussaint (2008a) and Thul (2008).

Table 1. Spearman rank correlations between nPVI and various complexity measures

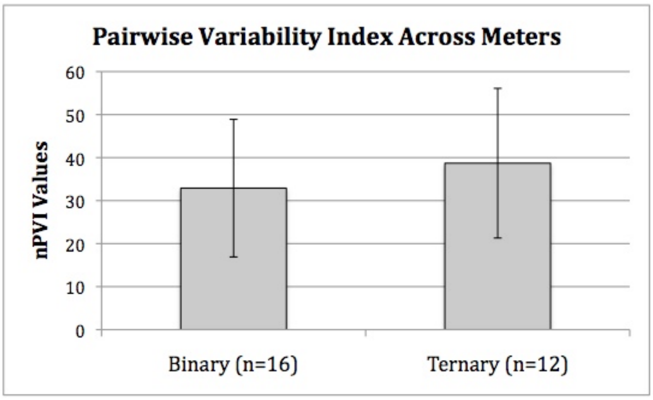
Complexity Measure	nPVI
Performance Complexity (Povel-Essens)	$r = -0.006$ $p < 0.5$
Perceptual Complexity (Povel-Essens)	$r = 0.100$ $p < 0.27$
Standard Deviation (Povel-Essens)	Stan. Dev. = 0
Performance Complexity (Essens)	$r = 0.047$ $p < 0.42$
Perceptual Complexity (Essens)	$r = 0.006$ $p < 0.49$
Standard Deviation (Essens)	**$r = 0.67$ $p < 0.01$**
Performance Complexity (Fitch-Rosenfeld)	**$r = 0.40$ $p < 0.02$**
Standard Deviation (Fitch-Rosenfeld)	**$r = 0.57$ $p < 0.01$**
Keith's Complexity Measure - C(S)	**$r = 0.65$ $p < 0.01$**
Stan. Dev. (12 North Indian Talas)	**$r = 0.73$ $p < 0.01$**
Stan. Dev. (30 South Indian Talas)	**$r = 0.67$ $p < 0.01$**
Stan. Dev. (116 Irregular Desi talas)	**$r = 0.77$ $p < 0.01$**
Stan. Dev. (14 African; k=5, n=12)	**$r = 0.80$ $p < 0.01$**
Stan. Dev. (16 African; k=5, n=16)	**$r = 0.86$ $p < 0.01$**
Stan. Dev. (39 African; k=4-9, n=12)	$r = 0.22$ $p < 0.09$
Stan. Dev. (37 African; k=3-10, n=16)	**$r = 0.56$ $p < 0.01$**
Stan. Dev. (46 Euclidean Rhythms)	**$r = 0.55$ $p < 0.01$**
Stan. Dev. (28 Random Rhythms; n=16)	**$r = 0.55$ $p < 0.01$**
Stan. Dev. (50 Rumanian Folk Rhythms)	**$r = 0.35$ $p < 0.01$**
Stan. Dev. (19 Arabian *Wazn*)	$r = 0.31$ $p < 0.1$
Stan. Dev. (20 Golomb Rulers)	$r = 0.31$ $p < 0.1$

The performance and perceptual complexity judgments in the Povel-Essens and Essens data sets do not correlate at all with the nPVI values. Furthermore, in the case of the Povel-Essens data, the nPVI cannot even be compared with the standard deviation because the standard deviation of the IOIs is the same for all the rhythms due to the fact that they all consist of permutations of the intervals in the set {1,1,1,1,1,2,2,3,4}. Thus the Spearman rank correlation coefficient is not even computable for this type of data. The Fitch-Rosenfeld rhythms tell a different story, however. Here, the human performance complexity is moderately but statistically significantly correlated with the nPVI ($r = 0.40$ with $p < 0.02$), as is the standard deviation ($r = 0.57$ with $p < 0.01$). As pointed out earlier, the Fitch-Rosenfeld rhythms differ from the other data sets in that seventeen of the thirty rhythms have a silent first pulse. In order to compute the nPVI the first IOI in the sequence was taken to be the silent interval. This is equivalent to making the first pulse of every rhythm an onset, and will be considered in more detail in the section describing the proposed refinement of the nPVI. For the Essens rhythms, the standard deviation is also highly statistically significantly correlated with the nPVI ($r = 0.67$, $p < 0.01$).

Discrimination Between Binary and Ternary Rhythms

Patel and Daniele (2003) compared the rhythm of English and French music and language using the nPVI of the lengths of the notes in music, and the vocalic durations in speech, respectively. In both cases they found that the nPVI values for British English were greater than for French. To test the influence of musical meter on the nPVI values of note durations and to account for the differences observed, they separated their musical themes into those with binary and ternary meters, and found that the nPVI values of the binary and ternary music corpora did not differ significantly. Their corpora consisted of 137 English musical themes from composers such as Elgar, Delius, and Holst, and 181 French musical themes from composers that included Debussy, Honegger, and Ravel. These results motivated the exploration of whether there is any significant difference between the nPVI values in a completely different musical context: African rhythm instead of classical European music, and note durations replaced by IOIs. For this purpose two groups of rhythms were compiled from a collection of books and journals consisting of thirty-seven binary 16-pulse timelines, and thirty-nine ternary 12-pulse timelines. The average nPVI values are shown in Figure 6, where the error bars indicate one standard deviation above and below the mean. To test whether the means are statistically significantly different the Student's *t*-test was used (Press et al. 1992, 616). The most notable aspect is that the two means differ little in absolute terms compared to the variability within each group, mirroring somewhat the findings by Patel and Daniele (2003). Indeed, the one-sided *t*-test yielded a value of $t = 1.542$, with $p = 0.0636$. On the other hand, the Kolmogorv-Smirnov test, which yields a distance value D, rejects the null hypothesis that the two distributions of nPVI values are the same, yielding a distance $D = 0.33$ with $p < 0.02$.

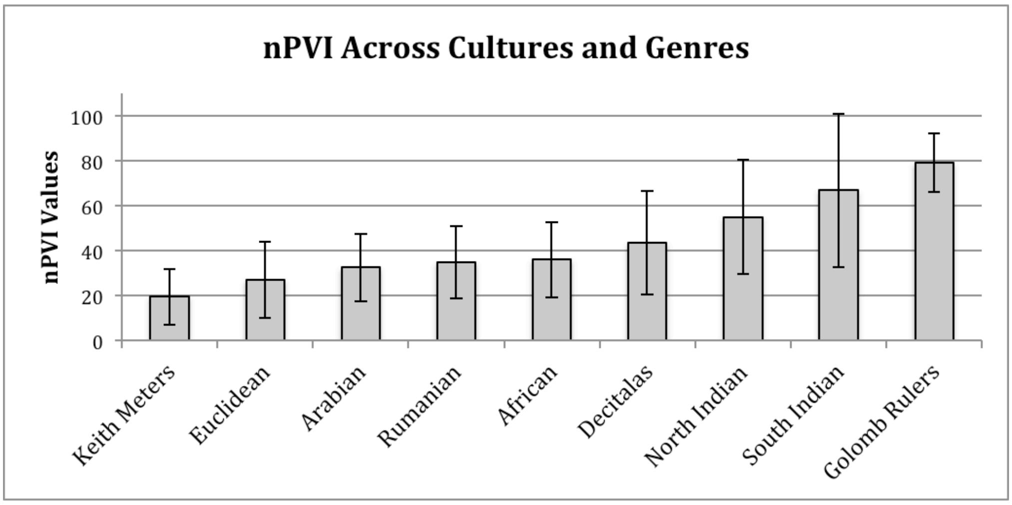
Figure 6. The nPVI values for the binary and ternary rhythm timelines common in the African diaspora



Discrimination Across Different Cultures and Genres

As mentioned in the introduction, it has been shown in previous research that the nPVI could be used to discriminate among different compositional styles such as nineteenth-century French and German art song (VanHandel 2006; VanHandel and Song 2009). These results motivated the testing of whether the nPVI is able to discriminate between musical rhythms of different genres and cultures, as well as non-musical "rhythms" such as Golomb rulers. To this end, several corpora of rhythms were compiled: commonly used meters, Euclidean rhythms, Arabian rhythms, Rumanian rhythms, African timelines, Indian talas, and Golomb rulers. The average values of the nPVI scores for nine of these corpora are shown in Figure 7 in order of increasing mean (average) nPVI values, along with error bars indicating plus and minus one standard deviation from the mean. Although for many pairs of families the variation within each group is much greater than the differences between their means, the general trend is that among musical rhythms the commonly used meters in Keith's list and the Euclidean rhythms are the simplest (most regular), and the Indian talas the most complex (irregular). However, the Golomb rulers on the far right top all musical rhythms in terms of the amount of complexity (irregularity) they contain. The notion of the complexity of rhythms, and its measurement, are themselves complex issues, and there exists a plethora of methods for measuring complexity (Thul 2008; Toussaint 2013: 107–120). The nPVI is but one such indicator.

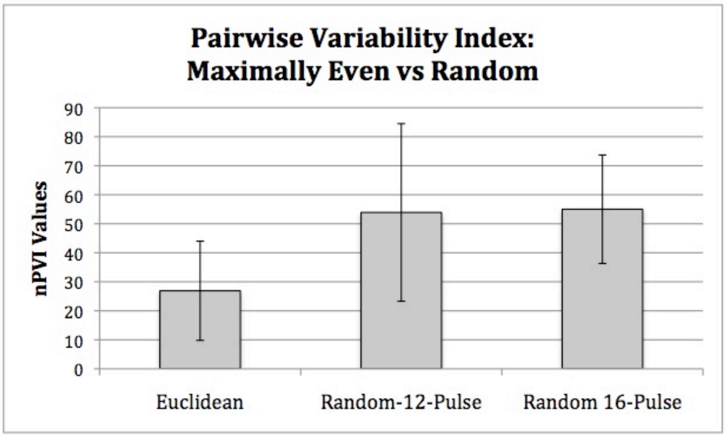
Figure 7. The average nPVI values for rhythms from different cultures and genres



Comparison with Random Rhythms

To obtain some insight into the sensitivity of the variability of the nPVI for different genres of rhythms the nPVI was also computed for two sets of random rhythms that were generated as described in Section III. The nPVI values for the binary 16-pulse and ternary 12-pulse rhythms are shown in Figure 8, alongside the value for the Euclidean (maximally even) rhythms. As might be expected, the nPVI values for the random rhythms are much higher than for the Euclidean rhythms. A Kolmogorov-Smirnov test of the difference between the distributions of the nPVI

Figure 8. Comparing Euclidean (maximally even) with random rhythms



values of the two families of rhythms yields a distance of $D = 0.65$ with $p < 0.001$, confirming the intuition that the maximally even rhythms are as a group highly non-random. However, the variability for the 16-pulse random rhythms is not greater than that for any of the other families of rhythms tested. Indeed, from Figure 7 it may be observed that the North Indian rhythms and desi talas have the highest variability of all. Does this mean that the Indian talas are more random than the rhythms of other genres? On the contrary, one might expect that variability is greater in specifically designed rhythms than in random rhythms. To test this conceivable hypothesis, Kolmogorov-Smirnov tests were performed with the nPVI values of the random rhythms (12-pulse and 16-pulse rhythms combined) and the three types of Indian talas separately: North Indian talas, South Indian sulaadi talas, and desi talas. The results are listed in Table 2. For the South Indian talas and desi talas we may reject the hypothesis that they come from the same distribution as the random rhythms. On the other hand, for the North Indian talas this is not the case, although this may be due to the small sample size of the North Indian talas. Comparing the nPVI values of the three systems of Indian talas with each other using Kolmogorov-Smirnov tests suggests that the desi

Table 2. Kolmogorov-Smirnov tests for nPVI values of the Indian talas and random rhythms

Kolmogorov-Smirnov Tests Against Random Rhythms		
North Indian Talas	South Indian Talas	Desi Talas
$D = 0.328$ $p < 0.21$	$D = 0.327$ $p < 0.03$	$D = 0.364$ $p < 0.001$

talas are significantly different from the North Indian talas ($D = 0.389$ with $p < 0.05$), as well as the South Indian talas ($D = 0.378$ with $p < 0.001$), but the North Indian talas are not significantly different from the South Indian talas ($D = 0.316$ with $p < 0.3$).

From Figure 7 it may be observed that the Arabian, Rumanian, and African rhythms have almost equal average nPVI values, as well as similar degrees of variation. Furthermore, all three genres of rhythms have similar short IOIs with a generous supply of IOIs made up of durations of two and three pulses. Therefore, these data sets provide strict tests of the power of the nPVI to discriminate between these genres of rhythms coming from different cultures. Table 3 shows the Kolmogorov-Smirnov distances and significance tests of the nPVI values obtained across cultures and genres for the nine families of rhythms of Figure 7. The results with statistically significant *p*-values (less than 0.05) are highlighted in bold. Of the thirty-six pairwise comparisons, thirty are statistically significant. The nPVI can easily distinguish between the North Indian talas and all the remaining data sets. Keith's meters (the most regular), and Golomb rulers (the most irregular) are significantly different from all the other groups. The Euclidean rhythms are significantly different from all other groups except the Arabian rhythms. The South Indian talas are significantly different from all other groups except the North Indian talas. The African timelines and Indian desi talas are difficult to distinguish from the Arabian rhythms at the 0.06 level. The families hardest to distinguish are the Arabian from the Rumanian and African rhythms, and the Rumanian from the African rhythms.

Table 3. Kolmogorov-Smirnov tests for nPVI values across cultures and genres

Kolmogorov-Smirnov Tests Across Cultures and Genres								
	Euclidean	Arabian	Rumanian	African	Desi Talas	North Indian	South Indian	Golomb Rulers
Keith Meters	$D = 0.29$ $p < 0.012$	$D = 0.44$ $p < 0.003$	$D = 0.48$ $p < 0.001$	$D = 0.47$ $p < 0.001$	$D = 0.61$ $p < 0.001$	$D = 0.75$ $p < 0.001$	$D = 0.75$ $p < 0.001$	$D = 1.0$ $p < 0.001$
Euclidean	-	$D = 0.23$ $p < 0.40$	$D = 0.28$ $p < 0.04$	$D = 0.35$ $p < 0.01$	$D = 0.38$ $p < 0.01$	$D = 0.59$ $p < 0.01$	$D = 0.56$ $p < 0.001$	$D = 0.89$ $p < 0.001$
Arabian		-	$D = 0.14$ $p < 0.94$	$D = 0.19$ $p < 0.60$	$D = 0.31$ $p < 0.06$	$D = 0.61$ $p < 0.01$	$D = 0.60$ $p < 0.001$	$D = 1.0$ $p < 0.001$
Rumanian			-	$D = 0.13$ $p < 0.66$	$D = 0.24$ $p < 0.03$	$D = 0.56$ $p < 0.01$	$D = 0.51$ $p < 0.001$	$D = 0.91$ $p < 0.001$
African				-	$D = 0.2$ $p < 0.05$	$D = 0.53$ $p < 0.004$	$D = 0.47$ $p < 0.001$	$D = 0.86$ $p < 0.001$
Desi Talas					-	$D = 0.39$ $p < 0.05$	$D = 0.36$ $p < 0.003$	$D = 0.72$ $p < 0.001$
North Indian						-	$D = 0.31$ $p < 0.295$	$D = 0.58$ $p < 0.007$
South Indian							-	$D = 0.41$ $p < 0.021$

A New Refinement of the nPVI.

It is an obvious fact that the order of the duration intervals of a rhythm influences the rhythm's perceived complexity by creating contrast between intervals and their adjacent intervals, as well as between intervals and the underlying meter. However, as already pointed out, the standard deviation of the IOIs is by definition blind to this order. Not surprisingly, it is not difficult to create examples that show that the standard deviation fails to completely characterize the complexity of rhythm timelines. Consider the well-known Cuban son clave and Brazilian bossa nova rhythm with IOI durational patterns [3-3-4-2-4] and [3-3-4-3-3], respectively (Toussaint 2002). The bossa nova rhythm is considered to be more complex (and syncopated) than the son clave, but the standard deviation of the latter is 0.837, whereas for the former it is only 0.447. Although the nPVI takes order into account and is thus contrast-sensitive, it is still oblivious to the underlying meter. Hence in this example the nPVI fares no better, yielding a value of 40.5 for the son clave and 14.3 for the bossa nova. Other equally anomalous examples are easy to find. The shiko [4-2-4-2-4] is much simpler than the son clave but its nPVI value of 66.7 is much greater.

Barry, Andreeva, and Koreman (2009) expose additional limitations of the nPVI in its ability to capture perceived rhythm in the Bulgarian, English, and German languages. These results support the thesis of Arvaniti (2009), who argues that metrics such as the standard deviation as well as the nPVI are unreliable predictors of rhythmic types in languages. Nolan and Asu conclude from their study that in language, duration cannot be "assumed to be either the exclusive correlate of perceived rhythm nor to act independently of other cues in perception" (2009, 75), and according to Royer and Garner, "pattern organizations are wholistic" (1970, 115). The results of the present study suggest similar conclusions with respect to short rhythms in the musical domain. Nevertheless, as the results in Table 3 attest, the nPVI is successful at discriminating between thirty of the thirty-six pairs of rhythm families tested.

To be more accurate and generally useful for both theoretical and practical musical rhythm analysis, a suitable modification of the nPVI that takes metrical information into account is desirable. What the exact nature of such a modification should entail in order to increase the correlation between the nPVI and human measures of complexity is as yet an open question. However, investigations along these lines have already begun. London and Jones (2011) proposed several modifications of the nPVI for its specific application to musical rhythm. These refinements include the application of the nPVI hierarchically to higher levels of rhythmic structure, the analysis of binary and ternary (duple and triple) rhythms separately, and the use of alternate codings of the IOI durations, such as rounding the durations to the nearest beat.

Here a new refinement of the nPVI is proposed that is motivated by the desire to modify the nPVI so that it incorporates metric information, and by the three human performance complexity measures investigated by Fitch and Rosenfeld (2007): (1) performance complexity, (2) beat-tapping complexity, and (3) number of resets. Here, "beat" refers to the 4/4 meter of the 16-pulse rhythms used in the experiments. The *performance* complexity is a measure of discrepancy between a rhythm heard, as played by a computer, and its subsequent reproduction by the listener by tapping. It is important to note that in these experiments the rhythm is first played together with the 4/4 beat (meter) provided by a low frequency "bass drum" sound, which is continuously played by the computer during the rhythm reproduction task. The *beat-tapping* (meter) complexity is a measure of discrepancy between the meter played together with a rhythm, and the subsequent reproduction of the meter by tapping, after the metric beat stops sounding. Similarly, in this test the rhythm continues to be played after the beat stops sounding. During the beat-tapping listening tests, the subjects sometimes alter their perception of where the metric beats are by shifting their position during the beat-tapping trial. This shift is termed a *reset*. More specifically, in the words of Fitch and Rosenfeld, "A reset event was scored when a subject's tap occurred closer in time to the syncopated pulse (the 'offbeat') than to the correct, unsyncopated pulse (the 'onbeat'), where the syncopated pulse falls exactly midway between unsyncopated pulses. After the data had been optimally aligned, the number of reset events as so defined was summed for each trial" (2007, 49). This *number of resets* determines the third measure of complexity.

In the experiments performed by Fitch and Rosenfeld (2007) the subjects first listened to the rhythm and meter together, then tapped each part individually while listening to the other part. Immediately preceding this test the subjects listened to and tapped either the rhythm or the meter while listening to both simultaneously. Therefore, the subjects were cognitively processing the resultant rhythms obtained from the union of the rhythms' onsets and the beats of the 4/4 meter. This observation motivated probing the Fitch-Rosenfeld data deeper with the following modification of the nPVI that incorporates the meter of the rhythm. The *Modified nPVI* (M-nPVI) of a rhythm *R*, heard in the context of a meter *M*, is defined as the standard nPVI computed on the resultant rhythm *RM* obtained from the union of *R* and *M*. For example, if the rhythm *R* is the son clave given by [x . . x . . x . . . x . x . . .] and the 4/4 meter *M* is [x . . . x . . . x . . . x . . .], then the resultant rhythm *RM* is [x . . x x . x . x . x . x . . .]. Thus it was felt that this measure, although quite different from the standard nPVI, was more suitable for testing in situations where a rhythm and its meter are felt together either acoustically or internally in the mind of the performer. Consider how the Modified nPVI compares with the standard nPVI for the four rhythms compared at the start of this section. They were, in order of increasing complexity (and syncopation), the shiko (nPVI = 66.7), the son (nPVI = 40.5), and the bossa nova (nPVI = 14.3). Note that the nPVI values are totally off the mark, producing a list in the reverse order of what is expected. The Modified nPVI values for these rhythms are: shiko (R-nPVI = 26.7), son (R-nPVI = 38.9), and bossa nova (R-nPVI = 47.6), in perfect agreement with musicological judgments.

The Spearman rank correlation coefficients were computed for the Modified nPVI with all thirty rhythms in the Fitch-Rosenfeld data. In addition, since seventeen of the thirty rhythms had silent first pulses, the correlations were also calculated

Table 4. Spearman Rank Correlations Between nPVI and Human Performance Complexities for Fitch-Rosenfeld Rhythms

Spearman Rank Correlations for nPVI with Fitch-Rosenfeld Rhythms			
	Performance	Beat-Tapping	Number of Resets
nPVI (Entire Data)	$\mathbf{r = 0.40}$ $\mathbf{p < 0.014}$	$r = -0.045$ $p < 0.41$	$r = 0.043$ $p < 0.41$
nPVI (Onset on First Pulse)	$r = 0.0$ $p < 0.5$	$r = -0.021$ $p < 0.24$	$r = 0.024$ $p < 0.47$
nPVI (Silent First Pulse)	$\mathbf{r = 0.68}$ $\mathbf{p < 0.002}$	$r = 0.20$ $p < 0.22$	$r = 0.21$ $p < 0.21$

Table 5. Spearman Rank Correlations Between Modified nPVI and Human Performance Complexities for Fitch-Rosenfeld Rhythms

Spearman Rank Correlations for Modified nPVI with Fitch-Rosenfeld Rhythms			
	Performance	Beat-Tapping	Number of Resets
Modified nPVI (Entire Data)	$r = 0.23$ $p < 0.11$	$\mathbf{r = 0.56}$ $\mathbf{p < 0.001}$	$\mathbf{r = 0.53}$ $\mathbf{p < 0.001}$
Modified nPVI (Onset on First Pulse)	$r = 0.42$ $p < 0.07$	$\mathbf{r = 0.71}$ $\mathbf{p < 0.003}$	$\mathbf{r = 0.77}$ $\mathbf{p < 0.001}$
Modified nPVI (Silent First Pulse)	$r = 0.21$ $p < 0.22$	$\mathbf{r = 0.59}$ $\mathbf{p < 0.007}$	$\mathbf{r = 0.52}$ $\mathbf{p < 0.017}$

independently with only the rhythms that contained an onset on their first pulse, and with only the rhythms that had a silent first pulse. The standard nPVI and Modified nPVI results are displayed in Tables 4 and 5, respectively, where the statistically significant values are highlighted in bold. For the standard nPVI (Table 4), only the performance complexity yielded statistically significant results. Using the entire data set a mild correlation with nPVI was observed ($r = 0.40$, $p < 0.014$), whereas using only the rhythms that contained a silent first pulse yielded a high correlation ($r = 0.68$, $p < 0.002$). The beat-tapping and number of resets measures showed no correlation with the nPVI. On the other hand, for the Modified nPVI (Table 5), the beat-tapping and number of resets measures are highly and statistically significantly correlated, especially for the rhythms that have an onset on their first pulse ($r = 0.71$, $p < 0.003$ and $r = 0.77$, $p < 0.001$, respectively).

Although the Modified nPVI gives improvements over the standard nPVI for the Fitch-Rosenfeld data, such improvements have limited generalizability. In particular, the two measures are effectively identical for rhythms that contain the metric beats as part of their onsets. Indeed, all the rhythms in the Essens data and almost all the rhythms in the Povel-Essens data have this property. Therefore, one cannot expect the Modified nPVI to offer improvements if the rhythms are not syncopated.

V. DISCUSSION

The principal goals of the research described here are to explore the efficacy of the nPVI as a tool for measuring the complexity of short musical rhythms, to compare it to the often used statistical measure of standard deviation, and to explore its possible applications to various problems that arise in musicology and music cognition. The main results exhibit two general trends that depend on the nature of data to which the nPVI is applied. The first type of data comprises predominantly human judgments of perceptual complexity obtained from listening tests, and human performance complexity evaluated with rhythms that were artificially algorithmically generated. The second kind of data consists of mathematical measures of complexity (nPVI, standard deviation, and Keith's metric complexity) that were computed on rhythms observed in the traditional musical practices of several cultures. On the artificially generated data the nPVI measure performed poorly, with one exception. On the traditional rhythms used in cultural practice it fared much better. While it is acknowledged that the use of artificially generated rhythm may be useful to test certain hypotheses, these results underscore the importance of also working with rhythms that are used outside of the laboratory.

The standard deviation has been used as a measure of variation for some time in language studies. More recently, research has moved towards using the nPVI in order to quantify the contrast between adjacent intervals, something that the standard deviation ignores completely. The results obtained here, however, show that the behavior of the two measures is, in general, not as different as one might hope. This is not to say that the nPVI does not offer substantial improvements over the standard deviation. Indeed, the Povel-Essens data set provides a compelling example of how totally useless the standard deviation of the IOIs can be for measuring the complexity of rhythms that are composed of only permutations of IOIs.

The correlation results in Table 1 exhibit a pattern of nPVI values that depend on the type of rhythms analyzed. The correlations between the complexity measures and the nPVI values tend to be either non-existent, low, or statistically insignificant for extreme rhythms or for those artificially generated by means of combinatorial methods, whereas the correlations are high and statistically significant for rhythms that are used in practice in traditional music. The highest correlation between the standard deviation and the nPVI of the IOIs is obtained with the sub-Saharan African timelines composed of five onsets: $r = 0.80$ with $p < 0.01$ for the ternary timelines, and $r = 0.86$ with $p < 0.01$ for the binary timelines.

The nPVI appears to be quite useful as a feature for discriminating between the rhythms of different cultures and genres, as the calculations in Figure 7 and Table 3 attest. Much has been written comparing the Indian talas to rhythms from other parts of the world, and within India, contrasting the rhythmic aspects of North Indian music with those of South India. Kuiper writes that “In North Indian music the talas are fewer and not organized in any systematic manner” (2010, 258). Furthermore, Kuiper adds that “North Indian talas have a further feature, the khali (‘empty’), a conscious negation of stress occurring at one or more points in each tala where one would expect a beat. There is nothing comparable to the khali in the South Indian system" (259). Hence, one might expect from these descriptions that North Indian talas might be different and perhaps more complex than South Indian talas. However, as the results of Table 3 indicate, the nPVI of the IOIs of the talas of these two genres of music fails to discriminate among them, suggesting that the difference between the two families of rhythms resides in aspects other than those dealing with the irregularity of the IOIs. Concerning the measurement of rhythmic complexity in Indian music Martin Clayton writes: "I can think of no objective criteria for judging the relative complexity or sophistication of rhythm in, for example Indian rag music, Western tonal art music, and that of African drum ensembles" (2000, 6). While this observation may at present be true enough with respect to the full richness of rhythm interacting at all its hierarchical levels, the nPVI does provide one objective measure for comparing meters, timelines, and talas at the surface level.

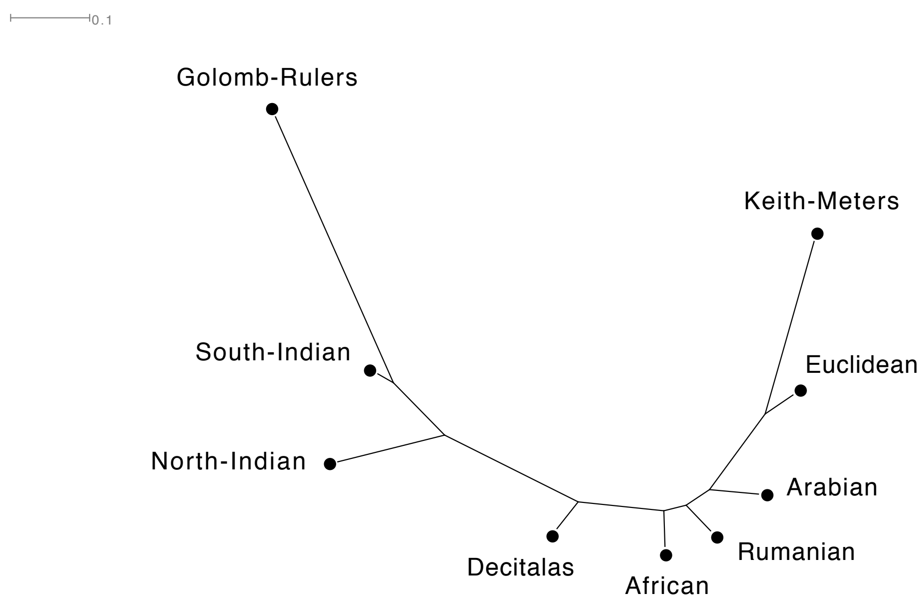
The complexity of Indian rhythm has also been compared to that of West Africa, some authors claiming each is more complex than the other. Kofi Agawu (1995) reviews a profusion of published claims about the alleged complexity of West African rhythms. With respect to Indian rhythms, the journalist and producer Joachim-Ernst Berendt repeats the exaggerated and doubtful Orientalist and Exoticist tropes: "It is necessary...to say a few words about the mysteries of Indian music. Its talas, its rhythmic sequences—incomprehensible for Western listeners—can be as long as 108 beats; yet the Indian ear is constantly aware of where the sam falls" (1987, 202). In his comparison of African and Indian music with European music Benjamin I. Gilman writes: "Hindu and African music is notably distinguished from our own by the greater complication of its rhythms. This often defies notation" (1909, 534). From the

Kolmogorov-Smirnov test results in Table 3 it may be concluded that Indian talas are quite different from African timelines according to the nPVI values of their IOIs. For North Indian talas $D = 0.53$, $p < 0.004$, and for South Indian talas $D = 0.47$, $p < 0.001$. Furthermore, from Figure 7 it may be surmised that Indian talas are much more complex than West African timelines when the nPVI of the IOIs is used as a measure of rhythm complexity. The average nPVI values for African timelines, and North Indian and South Indian talas are 35.13, 54.50, and 68.34, respectively. The *t*-tests for comparing the means for the North and South Indian talas with West African timelines yield values of 4.0 and 6.9, respectively, with *p*-values less than 0.0001, differences considered by conventional criteria to be extremely statistically significant.

The Arabian *wazn* (rhythmic cycles) and African timelines are in some respects quite different. Whereas the African timelines use cycles (measures) that are composed predominantly of twelve, sixteen, and twenty-four pulses, the Arabian wazn employ a much wider range of values that, like the Indian talas, may reach numbers as high as 176 (Touma 1996, 48). In that sense the Arabian wazn may be considered to be more complex than the African timelines. However, according to the nPVI they are not significantly different. The Kolmogorov-Smirnov test yields $D = 0.19$ with $p < 0.6$. This highlights the fact that the nPVI may be influenced more by the contrast in adjacent IOIs than by the number of pulses in the cycle, providing further evidence that the nPVI is insensitive to the underlying meter.

The relationships between all the Kolmogorov-Smirnov distances listed in Table 3 may be visualized at a glance by means of the *BioNJ* phylogenetic tree shown in Figure 9. Phylogenetic trees were originally developed in the context of evolutionary biology (Huson and Bryant 2006). The tree is used here for the purpose of visualization, and was computed using the *BioNJ* algorithm of Gascuel (1997), available in the *SplitsTree-4* software package (Dress, Huson, and Moulton 1996). The tree reveals three clusters, one comprising Golomb rulers, together with South and North Indian talas, a second consisting of the Keith meters and the Euclidean rhythms, and a third cluster that separates these two, made up of desi talas, African timelines, Rumanian rhythms, and Arabian wazn. The Kolmogorov-Smirnov distance between any two families of rhythms maps to the length of the shortest path between the corresponding nodes in the tree. Thus the furthest pair of rhythms consists of the Golomb rulers and the Keith meters. Since Golomb rulers make up the most complex rhythms, an alternate way to measure the complexity of a family of rhythms is by the magnitude of the Kolmogorov-Smirnov distance between the corresponding nodes in this tree; a small distance to the Golomb ruler node indicates a high degree of complexity. By this measure the tree in Figure 9 produces a clear ordering of the nine rhythm families in terms of decreasing complexity, as the tree is traversed from left to right. Furthermore, this ordering is identical to that obtained by ranking the nine families according to their average nPVI values illustrated in Figure 7.

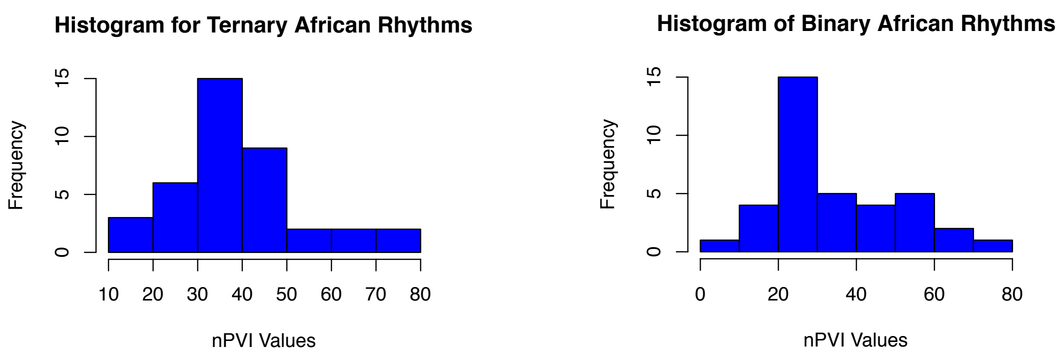
Figure 9. *BioNJ* phylogenetic tree of the Kolmogorov-Smirnov distances from Table 3.



The thirty rhythms collected by Fitch and Rosenfeld as well as their experimental results with three different measures of human performance complexity permitted several comparisons with the nPVI. Furthermore, the fact that about half of the rhythms had no onset on their first pulse permits dividing the data set into two classes to determine if any differences exist in the results for these two sub-families of rhythms. Two of the complexities, the beat-tapping complexity and the number of resets complexity, measure almost the same skill, namely how difficult it is to tap the meter while listening to the rhythm. On the other hand, the measure termed "performance complexity" quantifies how difficult it is to tap the rhythm while listening to the meter (Tables 4 and 5). The results of Table 4 indicate that the performance complexity involves a very different type of skill compared to the other two complexities. The beat-tapping and number of resets complexities are not correlated at all with the nPVI. These results also provide compelling evidence that the results differ radically for the two sub-families of rhythms when it comes to performance complexity. For the rhythms with a silent first pulse $r = 0.68$ with $p < 0.002$, and yet for rhythms with an onset on the first pulse there is no correlation at all with the nPVI. Naturally, for the entire data the results fall in between ($r = 0.40$ with $p < 0.014$). These results suggest that reproducing rhythms by tapping while listening to the meter is more difficult when the rhythms lack an onset on their first pulse. The results in Table 5 with the Modified nPVI are the converse of those in Table 4. Here the performance complexity is not correlated with the Modified nPVI for either the entire data or any of its two subsets, whereas all the other correlations are highly significant. These results, in combination with those of Table 4, suggest that the difficulty encountered by the subjects when tapping the meter while listening to the rhythm is determined by the irregularity of the resultant rhythm obtained from the union of the original rhythm and the meter.

Two fundamental classes of rhythms are the binary (duple) and ternary (triple). In African music these two classes are the most prevalent, with binary and ternary rhythm cycles being composed mostly of sixteen and twelve pulses, respectively. Christopher Hasty (1997, 132) argues that triple meter is markedly more complex than duple meter. Indeed, Hannon and Johnson (2004) and Huang et. al. (2012) have demonstrated experimentally that infants and adults, respectively, can discriminate easily between binary and ternary rhythms. On the other hand, as the results of Figure 6 illustrate, the average nPVI is unable to clearly discriminate between the binary and ternary African timelines (t = 1.542, with p = 0.0636, one-sided test). Nevertheless, the two distributions of the nPVI are quite different from each other, as the histograms in Figure 10 make plain. Indeed, as pointed out in the results section the Kolmogorov-Smirnov distance between the distributions is given by D = 0.33 with p < 0.02. Since the Modified nPVI correlates highly with the human beat-tapping and number of resets complexity measures (see Table 5), whereas the standard nPVI does not, one might wonder if the poor discrimination with the mean is due to the fact that it does not

Figure 10. Histograms of the nPVI values of the binary and ternary African rhythms.



incorporate metric information. However, the results with the Modified nPVI are even worse in this setting; the Kolmogorov-Smirnov test yields $D = 0.375$ with $p < 0.2$, and the *t*-test yields a value of 1.26 with $p < 0.1$. These results suggest that these complexity measures are not salient features that humans use to discriminate between duple and triple rhythms, and that the Modified nPVI has additional limitations.

VI. CONCLUSION AND FURTHER DIRECTIONS

The results obtained here indicate that the nPVI and the Modified nPVI are promising and powerful tools in certain contexts, although the precise general nature of these contexts has yet to be fully explored. The results also indicate that the nPVI of the IOIs, at least for some families of rhythms, does not differ markedly from the standard deviation of these intervals.

One observation that is common to all the results obtained here, evidenced from Figures 6, 7, and 8, is that the nPVI measure has a high variance. This sensitive behavior to outliers has been previously noted in the research on language (Wiget et al. 2010). To combat this sensitivity Jian (2004) proposed a modification of the original nPVI formula that incorporates the median rather than the mean of the adjacent interval differences. It has yet to be determined whether this variant of the nPVI would also improve the results in the music domain, such as those obtained here.

The marked difference in results obtained with rhythms that do or do not contain onsets on their first pulse provides compelling evidence suggesting that future studies of rhythm should treat the two categories of rhythms separately to better flesh out the effects of meter on other parameters.

Although the Modified version of the nPVI proposed here appears to be useful for some applications, the results also highlight its weaknesses in that it fails to discriminate between the binary and ternary African rhythms, as well as rhythms with onsets that contain the metric beats. It would be interesting to explore further refinements of the nPVI in the hope of obtaining a complexity measure that takes metric information into account in a more general and salient manner.

ACKNOWLEDGMENT

This research was supported by a research grant from the Provost's Office of New York University Abu Dhabi, Abu Dhabi, United Arab Emirates, administered through the Faculty of Science.

REFERENCES

Agawu, Kofi. 1995. *African Rhythm: A Northern Ewe Perspective*. Place: Cambridge University Press.

———. 2006. "Structural Analysis or Cultural Analysis? Competing Perspectives on the 'Standard Pattern' of West African Rhythm." *Journal of the American Musicological Society* 59(1): 1–46.

Aksentijevic, Aleksandar, and Keith Gibson. 2003. "Towards a Universal Measure of Structural Complexity." *Proceedings of the 8th International Conference of Cognitive Science*, Volume 1, edited by Peter Slezak, University of New South Wales, Sydney, 12–18.

———. 2012. "Complexity Equals Change." *Cognitive Systems Research* 15–16: 1–16. doi:10.1016/j.cogsys.2011.01.002.

Alperin, Roger C., and Vladimir Drobot. 2011. "Golomb Rulers." *Mathematics Magazine* 84: 48–55.

Amiot, Emmanuel. 2007. "David Lewin and Maximally Even Sets." *Journal of Mathematics and Music* 1(3): 157–72.

Arvaniti, Amalia. 2009. "Rhythm, Timing and the Timing of Rhythm." *Phonetica* 66: 46–63.

Asu, Eva L. and Francis Nolan. 2005. "Estonian Rhythm and the Pairwise Variability Index." *Proceedings of FONETIK*, Department of Linguistics, Göteborg, 29–32.

Barry, William B., and Jacques Koreman. 2009. "Do Rhythm Measures Reflect Perceived Rhythm?" *Phonetica* 66: 78–94.

Berendt, Joachim-Ersnt. 1987. *Nada Brahma: The World is Sound.* Translated by Helmut Bredigkeit. Rochester, Vermont: Destiny Books.

Cerulo, Karen A. 1988. "Analyzing Cultural Products: A New Method of Measurement." *Social Science Research* 17: 317–52.

Chaitin, Gregory J. 1974. "Information Theoretic Computational Complexity." *IEEE Transactions on Information Theory* 20: 10–15.

Chater, Nick. 1999. "The Search for Simplicity: A Fundamental Cognitive Principle?" *The Quarterly Journal of Experimental Psychology* 52A(2): 273–302.

Clayton, Martin. 2000. *Time in Indian Music*. Oxford: Oxford University Press.

Clough, John, and Jack Douthett, 1991. "Maximally Even Sets." *Journal of Music Theory* 35: 93–173.

Dalla Bella, Simone, and Isabelle Peretz. 2005. "Differentiation of Classical Music Requires Little Learning but Rhythm." *Cognition* 96(B): 65–78.

Daniele, Joseph R., and Aniruddh D. Patel. 2004. "The Interplay of Linguistic and Historical Influences on Musical Rhythm in Different Cultures." *Proceedings of the 8th International Conference on Music Perception and Cognition*, Evanston, Illinois (August): 759–62.

Douthett, Jack, and Richard Krantz. 2007. "Maximally Even Sets and Configurations: Common Threads in Mathematics, Physics, and Music." *Journal of Combinatorial Optimization* 14(4): 385–410.

Dress, Andreas, Daniel Huson, and Vincent Moulton. 1996. "Analysing and Visualizing Sequence and Distance Data Using SPLITSTREE." *Discrete Applied Mathematics* 71: 95–109.

Ellis, John, Frank Ruskey, Joe Sawada, and Jamie Simpson, 2003. "Euclidean Strings." *Theoretical Computer Science* 30(1): 321–40.

Erdős, Paul, and Pál Turán. 1941. "On a Problem of Sidon in Additive Number Theory and on Some Related Problems." *Journal of the London Mathematical Society* 16: 212–15.

Essens, Peter. 1995. "Structuring Temporal Sequences: Comparison of Models and Factors of Complexity." *Perception and Psychophysics* 57(4): 519–32.

Fitch, William T., and Andrew J. Rosenfeld. 2007. "Perception and Production of Syncopated Rhythms." *Music Perception* 25(1): 43–58.

Franklin, Phillip. 1956. "The Euclidean Algorithm." *The American Mathematical Monthly* 63(9): 663–64.

Freeman, D. Kent. 1997. "Maximizing Irregularity and the Golomb Ruler Problem." Internal Report. C.N.R.S. Luminy, Centre de Physique Theorique, Marseille, France.

Gascuel, Olivier. 1997. "*BIONJ*: an Improved Version of the *NJ* Algorithm Based on a Simple Model of Sequence Data." *Molecular Biology and Evolution* 14: 685–95.

Gibbon, Dafydd, and Ulrike Gut. 2001. "Measuring Speech Rhythm." *Eurospeech 2001 Scandinavia*: *7th European Conference on Speech Communication and Technology*, Aalborg, Denmark (September).

Gilman, Benjamin I. 1909. "The Science of Exotic Music." *Science* 30(772): 532–35.

Goyette, Jeremiah S. 2012. "The Z-Relation in Theory and Practice." Ph.D. diss., University of Rochester.

Grabe, Esther, and Ee Low. 2002. "Durational Variability in Speech and the Rhythm Class Hypothesis." In *Papers in Laboratory Phonology 7*, edited by Carlos Gussenhoven and Natasha Warner, 515–46. Cambridge: Cambridge University Press.

Hagoel, Kobi. 2003. *The Art of Middle Eastern Rhythm*. Kfar Sava, Israel: OR-TAV Music Publications.

Hannon, Erin E., and Scott P. Johnson. 2004. "Infants Use Meter to Categorize Rhythms and Melodies: Implications for Musical Structure Learning." *Cognitive Psychology* 50(4): 354–77.

Hasty, Christopher. 1997. *Meter as Rhythm.* New York and Oxford: Oxford University Press.

Huang, Juan, Darik Gamble, Kristine Sarnlertsophon, Xiaoqin Wang, and Steven Hsiao. 2012. "Feeling Music: Integration of Auditory and Tactile Inputs in Musical Meter Perception." *PLoS ONE* 7(10): e48496.

Huron, David. 2006. *Sweet Anticipation: Music and the Psychology of Expectation*, Cambridge: The MIT Press.

Huron, David, and Joy Ollen. 2003. "Agogic Contrast in French and English Themes: Further Support for Patel and Daniele 2003." *Music Perception* 21: 267–72.

Huson, Daniel H., and David Bryant. 2006. "Application of Phylogenetic Networks in Evolutionary Studies." *Molecular Biology and Evolution* 23(2): 254–67.

Jian, Hua-Li. 2004. "An Improved Pair-Wise Variability Index for Comparing the Timing Characteristics of Speech." *Proceedings of the 8th International Conference on Spoken Language Processing*, Jeju Island, Korea, October 4–8.

Johnson, Robert S. 1975. *Messiaen*. Berkeley: University of California Press.

Keith, Michael. 1991. *From Polychords to Pólya: Adventures in Musical Combinatorics*. Princeton: Vinculum Press.

Kivy, Peter. 1993. *The Fine Art of Repetition: Essays in the Philosophy of Music*. Cambridge, UK: Cambridge University Press.

Knuth, Donald E. 1998. *The Art of Computer Programming*. Vol. 2, 3rd edition. Reading, MA: Addison Wesley.

Kuiper, Kathleen, ed. 2010. *The Culture of India*. New York: The Rosen Publishing Group and Britannica Educational Publishing.

Lempel, Abraham, and Jacob Ziv, 1976. "On the Complexity of Finite Sequences." *IEEE Transactions on Information Theory* 22(1): 75–81.

London, Justin, and Katherine Jones. 2011. "Rhythmic Refinements to the nPVI Measure: A Reanalysis of Patel and Daniele (2003a)." *Music Perception* 29(1): 115–20.

London, Justin. 1995. "Some Examples of Complex Meters and Their Implications for Models of Metric Perception." *Music Perception* 19(4): 529–50.

Longuet-Higgins, H. Christopher, and Christopher Lee. 1984. "The Rhythmic Interpretation of Monophonic Music." *Music Perception* 1: 424–41.

McDonough, Joyce, Harold Danko, and Jason Zentz, 2007. "Rhythmic Structure of Music and Language: An Empirical Investigation of the Speech Cadence of American Jazz Masters Louis Armstrong and Jelly Roll Morton." *University of Rochester Working Papers in the Language Sciences* 3(1): 45–56.

McGowan, Rebecca W., and Andrea G. Levitt. 2011. "A Comparison of Rhythm in English Dialects and Music." *Music Perception* 28(3): 307–14.

Morris, Robert. 1998. "Sets, Scales, and Rhythmic Cycles: A Classification of Talas in Indian Music." In *21st Annual Meeting of the Society for Music Theory*, Chapel Hill, NC.

Nolan, Francis, and Eva L. Asu. 2009. "The Pairwise Variability Index and Coexisting Rhythms in Language." *Phonetica* 66: 64–67.

Olivo, Ann M. 2011. "Exploring the Speech Rhythm Continuum: Evidence from Ashanti Twi." *Journal of Speech Sciences* 1(3): 3–15.

Patel, Aniruddh D. 2008. *Music, Language, and the Brain*. New York: Oxford University Press.

Patel, Aniruddh D., and Joseph R. Daniele. 2003. "An Empirical Comparison of Rhythm in Language and Music." *Cognition* 87: B35–B45.

———. 2003. “Stress-Timed vs. Syllable-Timed Music? A Comment on Huron and Ollen (2003).” *Music Perception* 21(2): 273–76.

Povel, Dirk, and Peter Essens. 1985. “Perception of Temporal Patterns.” *Music Perception* 2: 411– 40.

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian, P. Flannery. 1992. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge, UK: Cambridge University Press.

Proca-Ciortea, Vera. 1969. “On Rhythm in Rumanian Folk Dance.” *Yearbook of the International Folk Music Council* 1: 176–99.

Psotka, Joseph. 1975. “Simplicity, Symmetry, and Syntely: Stimulus Measures of Binary Pattern Structure.” *Memory and Cognition* 3(4): 343–444.

Quinto, Lena, William F. Thompson, and Felicity L. Keating. 2013. “Emotional Communication in Speech and Music: The Role of Melodic and Rhythmic Contrasts.” *Frontiers in Psychology* 4(184): 1–8.

Rahn, Jay. 1987. “Asymmetrical Ostinatos in Sub-Saharan Music: Time, Pitch, and Cycles Reconsidered.” *In Theory Only* 9(7): 23–37.

———. 1996. “Turning the Analysis Around: African-Derived Rhythms and Europe-Derived Music Theory.” *Black Music Research Journal* 16(1): 71–89.

Raju, Marju, Eva L. Asu, and Jaan Ross. 2010. “Comparison of Rhythm in Musical Scores and Performances as Measured with the Pairwise Variability Index.” *Musicae Scientiae* 14(1): 51–71.

Ramus, Franck, Marina Nespor, and Jacques Mehler. 1999. “Correlates of Linguistic Rhythm in the Speech Signal.” *Cognition* 73: 265–92.

Rickard, Scott. 2011. The World's Ugliest Music, *TEDxTalks*. http://www.youtube.com/watch?v=RENk9PK06AQ

Royer, Fred L., and Wendell R. Garner. 1970. “Perceptual Organization of Nine-Element Auditory Temporal Patterns.” *Perception and Psychophysics* 7(2): 115–20.

Sadakaka, Makiko, Peter Desain, Henkjan Honing, Patel, Aniruddh D., and John R. Iversen. 2004. “A Cross-cultural Study of the Rhythm in English and Japanese Popular Music.” *Proceedings of the International Symposium on Musical Acoustics*, March 31 to April 3, Nara, Japan, 41–44.

Sandness, Frode E., and Hua-Li Jian. 2004. “Pair-Wise Variability Index: Evaluating the Cognitive Difficulty of Using Mobile Text Entry Systems.” In *Mobile Human-Computer Interaction: Proceedings of the 6th International Symposium*, edited by Stephen Brewster and Mark Dunlop, 347–50. Berlin: Springer.

Shearer, James B. 2012. Golomb Rulers. http://www.research.ibm.com/people/s/shearer/gropt.html

Shmulevich, Ilya, and Dirk-Jan Povel. 2000. “Measures of Temporal Pattern Complexity.” *Journal of New Music Research* 29(1): 61–69.

Smith, Leigh, and Henkjan Honing. 2006. “Evaluating and Extending Computational Models of Rhythmic Syncopation in Music.” *Proceedings of the International Computer Music Conference*, 688–91.

Soderberg, Stephen. 2013. A Note on Golomb Rulers and Z-Related Sets, *Academia.edu*, January 26. http://www.academia.edu/2466601/A_Note_on_Golomb_Rulers_and_Z-Related_Sets.

Spearman, Charles. 1904. “The Proof and Measurement of Association Between Two Things.” *American Journal of Psychology* 15: 72–101.

Standifer, James A. 1988. “The Tuareg: Their Music and Dances.” *The Black Perspective in Music* 16(1): 45–62.

Thul, Eric. 2008. “Measuring the Complexity of Musical Rhythm.” M.Sc. Thesis, McGill University.

Thul, Eric, and Godfried T. Toussaint. 2008a. “Rhythm Complexity Measures: A Comparison of Mathematical Models of Human Perception and Performance.” *Proceedings of the 9th International Conference on Music Information Retrieval,* Philadelphia, September 14–18, 663–68.

———. 2008b. “A Comparative Phylogenetic Analysis of African Timelines and North Indian Talas.” *Proceedings of BRIDGES: Mathematics, Music, Art, Architecture, and Culture*.

Touma, Habib H. 1996. *The Music of the Arabs*. Portland, Oregon: Amadeus Press.

Toussaint, Godfried T. 2002. “A mathematical analysis of African, Brazilian, and Cuban clave rhythms.” *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science,* Townson University, Towson, MD, July 27–29, 157–68.

———. 2005. “The Euclidean Algorithm Generates Traditional Musical Rhythms.” *Proceedings of BRIDGES: Mathematical Connections in Art, Music, and Science*, Banff, Alberta, Canada. 47–56.

———. 2011. “The Rhythm that Conquered the World: What Makes a "Good" Rhythm Good?” *Percussive Notes* (November): 52–59.

———. 2013. *The Geometry of Musical Rhythm.* Boca Raton, Florida: Chapman and Hall/CRC Press.

VanHandel, Leigh, 2006. "Trends in/over Time: Rhythm in Speech and Musical Melody in 19th-Century Art Song." *Proceedings of the 9th International Conference on Music Perception and Cognition*, University of Bologna, August, 1153–58.

VanHandel, Leigh, and Tian Song. 2009. "Influence of Linguistic Rhythm on Individual Compositional Style in 19th Century French and German Art Song." *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music*, Jyväskylä, Finland, 553–57.

Wiget, Lukas, Laurence White, Barbara Schuppler, Izabelle Grenon, Olesya Rauch, and Sven L. Mattys. 2010. "How Stable are Acoustic Metrics of Contrastive Speech Rhythm?" *Journal of the Acoustical Society of America* 127(3): 1559–69.

Yoon, Tae-Jin. 2010. "Capturing Inter-speaker Invariance Using Statistical Measures of Rhythm." *Proceedings of the Speech Prosody Conference*, Chicago, 1–4.