

## Measuring Musical Rhythm Similarity: Statistical Features Versus Transformation Methods

Juan F. Beltran

*Computer Science Department, New York University Abu Dhabi  
Abu Dhabi, P. O. Box 129188, United Arab Emirates  
[jfb325@nyu.edu](mailto:jfb325@nyu.edu)*

Xiaohua Liu

*Mathematics Department, New York University Abu Dhabi  
Abu Dhabi, P. O. Box 129188, United Arab Emirates  
[xl450@nyu.edu](mailto:xl450@nyu.edu)*

Nishant Mohanchandra\* and Godfried T. Toussaint<sup>†</sup>

*Computer Science Department, New York University Abu Dhabi  
Abu Dhabi, P. O. Box 129188, United Arab Emirates  
<sup>\*</sup>[nm1345@nyu.edu](mailto:nm1345@nyu.edu)  
<sup>†</sup>[gt42@nyu.edu](mailto:gt42@nyu.edu)*

Received 6 September 2013

Accepted 7 October 2014

Published 31 December 2014

Two approaches to measuring the similarity between symbolically notated musical rhythms are compared with each other and with human judgments of perceived similarity. The first is the edit-distance, a popular transformation method, applied to the symbolic rhythm sequences. The second approach employs the histograms of the inter-onset-intervals (IOIs) calculated from the rhythms. Furthermore, two methods for dealing with the histograms are also compared. The first utilizes the Mallows distance, a transformation method akin to the Earth-Movers distance popular in computer vision, and the second extracts a group of standard statistical features, used in music information retrieval, from the IOI-histograms. The measures are compared using four contrastive musical rhythm data sets by means of statistical Mantel tests that compute correlation coefficients between the various dissimilarity matrices. The results provide evidence from the aural domain, that transformation methods such as the edit distance are superior to feature-based methods for predicting human judgments of similarity. The evidence also supports the hypothesis that IOI-histogram-based methods are better than music-theoretical structural features computed from the rhythms themselves, provided that the rhythms do not share identical IOI histograms.

<sup>†</sup>Corresponding author.

*Keywords:* Musical rhythm; similarity measures; transformations; inter-onset interval histograms; mallows distance; edit distance; music information retrieval; symbolic music notation; Mantel test.

## 1. Introduction

One of the most fundamental problems in many scientific domains including cognition, perception, pattern recognition, computer vision, information retrieval, neuroscience and artificial intelligence is the measurement of similarity between a pair of objects.<sup>29–31,35</sup> There are two general approaches to tackling this problem that have received much attention in the literature: *feature-based* methods and *transformation-based* procedures. In feature-based methods a collection of  $d$  features (measurements) is calculated for each object, thus representing the object as a vector in a  $d$ -dimensional feature space. The dissimilarity of two objects is then characterized by some distance measure calculated between their corresponding feature vectors.<sup>7,18,28</sup> The transformation-based techniques on the other hand measure similarity between two objects by the minimum amount of *work* (suitably defined) that is required to *transform* (or morph) one object into the other.<sup>11,12</sup> When the objects to be compared are sequences of symbols, as is the case here with musical rhythms, transformation methods have been applied in a variety of domains such as text processing,<sup>15</sup> bioinformatics,<sup>23</sup> and music information retrieval.<sup>36</sup> Experimental evidence has been accumulating, for at least a decade, suggesting that in the visual perception domain, the transformation methods appear to be superior to the feature-based methods.<sup>5,6</sup> One such procedure applied to many pattern recognition problems is the *edit* distance (also known as the Levenshtein distance). Previous experiments have suggested that the edit distance is a good predictor of human perceptual judgments of rhythm similarity.<sup>34</sup> Furthermore, a recent comparison of the edit distance with a feature-based measure that employed a collection of *structural* features commonly used in music theory and ethnomusicology for the purpose of comparative analysis and classification, has highlighted the superiority of the edit distance, and has thus added support from the auditory domain to the growing consensus on the advantages of transformation methods already observed in the visual domain.

Here two approaches for measuring the similarity between symbolically notated musical rhythms are compared with each other and with human judgments of perceived similarity: the feature-based approach and the transformation method. In the feature-based approach, in contrast to the music-theoretical *structural* features used in a previous study,<sup>34</sup> the features employed here are *statistical* features (such as the mean, variance, kurtosis and other statistical moments) computed from the inter-onset interval (IOI) histograms. Contrary to what was done by previous researchers with acoustic input in music information retrieval applications,<sup>9,23</sup> here the IOI histograms are computed from the symbolically notated rhythms (as in Western music scores). For the transformation approach, two additional methods were

investigated. The first calculates the edit-distance directly on the symbolically notated rhythm sequences, and the second calculates the Mallows distance, a special case of the Earth-Movers distance, from the IOI-histograms of these rhythm sequences.<sup>14,36</sup> In order to determine the robustness of our methods the comparisons are made using four widely different musical rhythm data sets obtained from previous studies available in the literature. Some of the results and conclusions described here were previously summarized in a three-page conference proceedings extended abstract.<sup>3</sup> That extended abstract does not contain the definitions and descriptions of the Mallows distance measure, nor how it was generalized to apply to variable length rhythms, or the statistical features calculated. It only describes experiments performed with the first two data sets. Here we report on experiments with two additional data sets, we include phylogenetic trees computed for the Mallows distances, and provide a much broader and deeper discussion of the results.

The perception of rhythm in a general setting in which the input is a piece of music is a formidably complex process. Many parameters of music including intensity, tempo, pitch, and timbre affect the perception of rhythm. Even when the input is restricted to percussive sounds, as in this study, accents greatly influence this perception.<sup>22</sup> Arom has dissected the many different forms of accents possible in percussive rhythms.<sup>1,2</sup> See also the work of Rivière.<sup>21</sup> In order to simplify our study we assume the rhythms have no accents whatsoever. This not only makes sense scientifically in order to obtain more precise results, but it is also a common practice in Sub-Saharan African drum music in which rhythm timelines typically played with bells do not have accents.<sup>31</sup> Nevertheless, the methodology employed here may be applied also to notes produced by tone instruments. In such a situation the IOIs only code the beginnings of the notes (hence the term *onsets*). In many applications it is sufficient to consider only the onsets of notes.<sup>26</sup>

In some applications such as music information retrieval (MIR), features of the entire set of IOIs have been used to characterize the rhythms being analyzed, sometimes with the misguided intuition that the resulting (so-called full interval) histograms contain more information than the sparser adjacent interval histograms. However, full-interval IOI's in fact discard useful discrimination information in some cases, even if the rhythms being compared are not rotations or mirror images of each other, and thus sound very different from each other.<sup>19,25</sup> Indeed, even rotations and mirror images of a rhythm may sound quite different from each other, and clearly such rhythms have identical IOI histograms.<sup>31</sup> In the studies presented three of the four data sets (Tables 2–4) contain a varying number of rhythms with identical IOI histograms, which permits the determination of how robust our methods are when such IOI's are present in the data.

## 2. Transformation Methods

It is quite common to find examples in the music information retrieval literature, and pattern recognition in general, where similarity is measured either with features

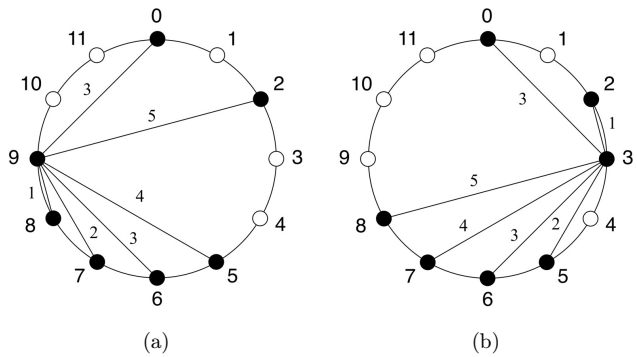


Fig. 1. Two different rhythms may have the same IOI histograms. The IOIs from the seventh onset (a) and the third onset (b) are highlighted with edges.

calculated from the IOI histograms, or by comparing histograms directly with each other.<sup>5,6,16,27</sup> However, less well known is the fact that two different sequences can have identical IOI histograms. Such sequences are called *homometric cyclotomic sets*, and have been extensively studied in the crystallography literature.<sup>8</sup> Consider the two rhythms shown in Fig. 1 in cyclic notation. Cyclic notation expresses a periodic rhythm as a circular lattice of pulses (points), which are black or white indicating, respectively, if a pulse is sounded or silent. A sounded pulse corresponds to an onset. The rhythm starts at the pulse marked “0” and time flows in a clockwise direction. The first rhythm in Fig. 1 (a) has *adjacent interval vector* [2-3-1-1-1-3], whereas the second (b) has adjacent interval vector [2-1-2-1-1-4]. These vectors are lists of the intervals between the pairs of adjacent onsets. Figure 1 shows an example discovered by Fichtner.<sup>8</sup> Clearly the two rhythms are neither rotations nor mirror image reflections of each other. This follows obviously from the fact that the rhythm on the right has an interval of duration four, which the rhythm on the left is lacking. Furthermore, the rhythm on the left has mirror symmetry (about the line through pulses 1 and 7), whereas the rhythm on the right has no mirror symmetry. Note that the rhythm on the right may be obtained from the one on the left by moving the onset at pulse 9 to the position of pulse 3, while preserving the six distances in question. Since the remaining distances are not affected by this operation, both rhythms have the same *full* IOI histogram shown in Fig. 2. The full IOI histograms contain the distances between *all* pairs of intervals (not just adjacent intervals). Note

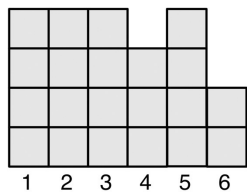


Fig. 2. The common full IOI histogram of the two rhythms shown in Fig. 1.

that these distances are *geodesics*, i.e. shortest paths along the circle. This example shows that information may be lost when comparing the histograms of rhythms rather than the rhythms themselves. One of the goals of this research is to determine how much discrimination information contained in the rhythms is lost in this abstraction process.

### 2.1. The Mallows distance

The Mallows distance<sup>14</sup> is closely related to the earth-mover's and transportation distances,<sup>36</sup> and is quite complicated to compute in its most general setting. In this study it is used in the context of binary rhythm sequences that may have different numbers of pulses and onsets. When rhythms have the same number of pulses and onsets the histograms of the IOIs have the same number of cells (total number of bin entries). In this special case, the Mallows distance simplifies greatly and may be computed efficiently in  $O(d)$  time, where  $d$  is the cardinality of the two histograms being compared, as follows. Let  $A$  and  $B$  be two rhythms with the same numbers  $n$  and  $k$ , respectively, of pulses and onsets, and let  $H_A$  and  $H_B$  denote their corresponding IOI histograms. Let each histogram consist of  $h$  bins containing a total of  $d$  elements, and let  $a_i$  and  $b_i$  denote the elements of  $H_A$  and  $H_B$ , respectively, sorted in lexicographical order. Furthermore, let  $x(a_i)$  and  $x(b_i)$  denote the “ $x$ -coordinates” (bin indices) of the elements, and refer to the example in Fig. 3, where  $h = 6$  and  $d = 8$ .

The sorted elements of the histograms determine the vectors:  $V_A = [x(a_1), x(a_2), \dots, x(a_d)]$  and  $V_B = [x(b_1), x(b_2), \dots, x(b_d)]$ , where  $x(a_i) \leq x(a_{i+1})$  and  $x(b_i) \leq x(b_{i+1})$  for  $i = 1, 2, \dots, d - 1$ . Then, the Mallows distance between the two histograms  $H_A$  and  $H_B$  is given by:

$$d_M(H_A, H_B) = \sum_{i=1}^d |x(a_i) - x(b_i)|. \quad (1)$$

This distance is the smallest total  $x$ -distance that all the elements of  $H_A$  must travel so that they end up in the positions of the elements of  $H_B$ . In the example of Fig. 3,  $V_A = (1, 2, 2, 3, 3, 3, 3, 5)$ ,  $V_B = (1, 1, 3, 3, 5, 5, 6, 6)$ , and the resulting Mallows distance is:  $0 + 1 + 1 + 0 + 2 + 2 + 3 + 1 = 10$ . By contrast, the standard bin-to-bin distance between the histograms is:  $1 + 2 + 2 + 0 + 1 + 2 = 8$ .

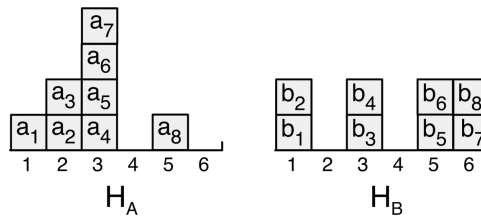


Fig. 3. The Mallows distance assignments between the elements of two histograms.

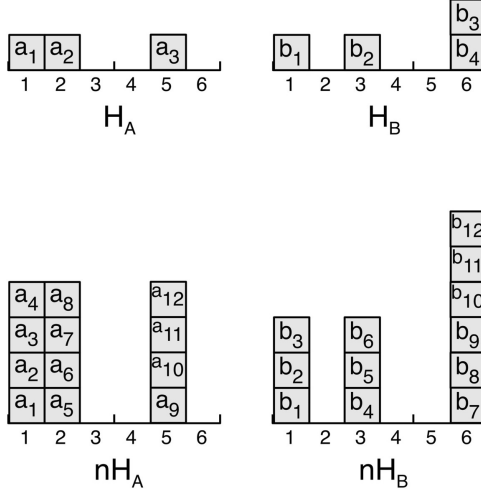


Fig. 4. The generalized Mallows distance assignments between the normalized histogram elements of two histograms  $H_A$  and  $H_B$  that contain an unequal number of cells.

In order to be able to compare rhythms that have the same number  $n$  of pulses, but a different number of onsets, the Mallows distance is generalized as follows. Let  $A$  be a rhythm with  $n$  pulses and  $k_A$  onsets, and let  $B$  be a rhythm with  $n$  pulses and  $k_B$  onsets. Let  $H_A$  and  $H_B$  denote their corresponding IOI histograms containing, respectively,  $d_A$  and  $d_B$  elements, and let  $a_i$  and  $b_i$  denote the elements of  $H_A$  and  $H_B$ , respectively, sorted in lexicographical order. Furthermore, as before, let  $x(a_i)$  and  $x(b_i)$  denote the “ $x$ -coordinates” (bin indices) of the elements, and refer to the example in Fig. 4 (top), where  $h = 6$ ,  $d_A = 3$  and  $d_B = 4$ . In order to have a one-to-one correspondence for “moving” all the elements of  $H_A$  to the positions of those of  $H_B$  the histograms are normalized so that they both contain the same number of elements. This is done by multiplying the heights of the histogram bins in  $H_A$  and  $H_B$  by two constants  $c_A$  and  $c_B$ , respectively, such that the product  $H_A \cdot c_A = H_B \cdot c_B = D$ , where  $D$  is the least common multiple (LCM) of  $d_A$  and  $d_B$ . In the example of Fig. 4  $d_A = 3$  and  $d_B = 4$ , thus  $c_A = 4$  and  $c_B = 3$  yielding a value of  $D = 12$ . Since this number is now the same for both histograms being compared, the generalized Mallows distance may be computed in a manner similar to that of the original version expressed by Eq. (1). More precisely, the generalized Mallows distance denoted by  $d_{GM}$ , is normalized by dividing by  $D$ , yielding Eq. (2). Note that if all the rhythms in the corpus are compared then the value of  $D$  becomes the LCM of the cardinalities of the elements contained in the histograms of all the rhythms in the corpus.

$$d_{GM}(nH_A, nH_B) = \frac{1}{D} \sum_{i=1}^D |x(a_i) - x(b_i)|. \quad (2)$$

In the example shown in Fig. 4 (bottom),  $V_{nHA} = (1, 1, 1, 1, 2, 2, 2, 2, 5, 5, 5, 5)$  and  $V_{nHB} = (1, 1, 1, 3, 3, 3, 6, 6, 6, 6, 6, 6)$ . The resulting generalized Mallows distance is thus:  $0 + 0 + 0 + 2 + 1 + 1 + 4 + 4 + 1 + 1 + 1 + 1 = 16/12$ , which yields  $d_{GM} = 4/3$ .

## 2.2. The edit distance

The edit distance measures the least amount of work needed to convert one rhythm to another by the minimum number of basic operations (*insertions*, *deletions*, and *substitutions* of symbols) that are sufficient to accomplish the task. An insertion of a symbol into a sequence lengthens the sequence by one symbol, a deletion shortens the sequence accordingly, and a substitution exchanges one symbol for another without altering the sequence's length. Consider the two eight-pulse binary sequences defined by  $A = [\times - \times - - \times - -]$  and  $B = [\times - - \times - \times - -]$ , where ' $\times$ ' denotes an onset (sounded pulse) and '-' denotes a rest (silent pulse). Here  $A$  may be converted into  $B$  by inserting a silent pulse after the first onset, and deleting a silent pulse after the second onset, for a total of two operations: one insertion and one deletion. Alternately, the second onset may be substituted by a silent pulse, after which the fourth pulse may be replaced with an onset, for a total of two substitutions. In general each operation may carry a different cost. We assume in this work that all three operations incur one unit of cost, and therefore both strategies in this example yield an edit distance equal to 2. A noteworthy advantage of the edit distance is that it may be used to compare rhythms that have different numbers of both onsets and pulses. Furthermore, previous experiments have provided evidence that the edit distance agrees well with human judgments of rhythm dissimilarity.<sup>20</sup> The edit-distance may be computed using dynamic programming in  $O(mn)$  worst-case time, where  $m$  and  $n$  are the cardinalities of the lengths of the two rhythm sequences.<sup>15</sup>

## 3. Structural Features

In a previous study, the edit distance was compared to a feature-based method that used a group of 14, structural features computed on the time spans of the rhythms themselves, that are frequently employed for rhythm analysis and classification in music theory and ethnomusicology.<sup>33</sup> The 14 features used were: (1) the parity of the number of pulses, (2) the parity of the number of onsets, (3) the density of the number of onsets (the number of onsets greater than one half the number of pulses), (4) anacrusis (the first pulse is silent), (5) the evenness of onset distribution, (6) vertical mirror symmetry, (7) horizontal mirror symmetry, (8) positively skewed mirror symmetry, (9) negatively skewed mirror symmetry, (10) rhythmic oddity measure of syncopation, (11) off-beatness measure of syncopation, (12) the hemiola property, (13) the toggle property, and (14) the balanced onsets property. The data used in that study consisted of the nine 8-pulse rhythms shown in Table 1. Thus each rhythm was converted into a 14-dimensional feature vector, and the relative

Table 1. The six Afro-Cuban 5-onset, 16-pulse rhythms.

| Rhythm Name | Binary Box Notation             |
|-------------|---------------------------------|
| Shiko       | × - - - × - × - - - × - × - - - |
| Son         | × - - × - - × - - - × - × - - - |
| Soukous     | × - - × - - × - - - × × - - - - |
| Rumba       | × - - × - - - × - - × - × - - - |
| Bossa-Nova  | × - - × - - × - - - × - - × - - |
| Gahu        | × - - × - - × - - - × - - - × - |

dissimilarity between two rhythms was taken to be the 1st-order Minkowski metric distance (or  $L_1$  norm) between their feature vectors. Listening tests with human subjects led to the conclusion that the edit distance was successful at predicting human judgments, whereas the feature-based method fared quite poorly.<sup>33</sup> One of the goals of this research is to determine if the statistical features computed on the IOI histograms, that have been used successfully in music information retrieval applications, are better than the structural features used in music theory, for predicting human perception of rhythm similarity.

#### 4. Statistical Features

The statistical features used in this study included the eight features calculated from the IOI histograms previously investigated by Scheirer *et al.*,<sup>24</sup> and Gouyon *et al.*<sup>9</sup> In the present study five additional features were included in the group, which were calculated from the IOIs themselves. These consisted of the shortest IOI, the longest IOI, the range of IOI values, the standard deviation of the IOIs, and the Normalized Pairwise Variability index (nPVI) of the IOIs. The definitions of these features are as follows:

- (1) Arithmetic mean:

$$AM = \frac{1}{N} \sum_{i=1}^N x_i. \quad (3)$$

- (2) Geometric mean:

$$GM = \left( \prod_{i=1}^N x_i \right)^{1/N}. \quad (4)$$

- (3) Total energy:

$$TE = \sum_{i=1}^N (x_i)^2. \quad (5)$$



(4) Centroid:

$$C = \frac{\sum_{i=1}^N i * x_i}{\sum_{i=1}^N x_i}. \quad (6)$$

(5) Flatness:

$$F = \log(\text{GM}) - \log(\text{AM}). \quad (7)$$

(6) Kurtosis:

$$K = \frac{\mu_4}{\mu_2^2} - 3. \quad (8)$$

(7) High frequency content:

$$\text{HFC} = \sum_{i=1}^N i * x_i^2. \quad (9)$$

(8) Skewness:

$$K = \frac{\mu_3}{\mu_2^{3/2}}. \quad (10)$$

(9) Longest IOI:

$$\text{Max} = \max_i \{x_i\}. \quad (11)$$

(10) Shortest IOI:

$$\text{Min} = \min_i \{x_i\}. \quad (12)$$

(11) Range of IOIs:

$$R = \max_i \{x_i\} - \min_i \{x_i\}. \quad (13)$$

(12) Standard deviation of IOIs:

$$\text{SD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \text{AM})^2}. \quad (14)$$

(13) Variability index (nPVI) of IOIs:

$$\text{nPVI} = \left( \frac{100}{m-1} \right) \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right|. \quad (15)$$

The nPVI measures the directional change in a sequence.<sup>10</sup> Unlike the standard deviation, which treats IOIs in isolation, the nPVI measures the deviations between adjacent IOIs. Here  $m$  is the number of IOIs in a rhythm, and  $d_k$  is the  $k$ th IOI in the sequence. The nPVI was originally proposed for the analysis of variability in speech rhythms using vocalic lengths. More recently it has been explored as a general tool in musical rhythm analysis.<sup>32</sup>

Table 2. The nine Afro-Cuban variable-onset 8-pulse rhythms.

| Rhythm Name       | Binary Box Notation |
|-------------------|---------------------|
| 2-3-3             | × − × − − × − −     |
| 3-2-3             | × − − × − × − −     |
| Cinquillo-Variant | × − × × − × × ×     |
| Cinquillo         | × − × × − × × −     |
| Conga             | × − − × − − − −     |
| Contradanza       | × − × × × × × −     |
| Habanera          | × − − × × − × −     |
| Tango-Congo       | × − − × × − − −     |
| Tresillo          | × − − × − − × −     |

## 5. Rhythm Data Sets

The calculations were carried out using four data sets that had been used in previous rhythm similarity tests with human subjects.<sup>20,33,34</sup> The first data set consists of six 16-pulse, 5-onset Afro-Cuban rhythm timelines shown in box notation in Table 1. These six rhythms play a distinguished role in much popular music of today.<sup>31</sup> In box notation the symbol “×” stands for a sounded unit time pulse (also called onset), and the symbol “−” stands for a silent unit time pulse.

The second data set consists of nine Afro-Cuban rhythms composed of eight pulses each with the number of onsets varying between two for the conga and six for the contradanza and cinquillo-variant, shown in box notation in Table 2.

The third data set consists of 17 rhythms composed of four, five, or six onsets and 16 pulses each, shown in box notation in Table 3. All these rhythms are *mutations* of the clave son rhythm. Sixteen of these rhythms were obtained by mutating every one of the 16 pulses of the clave son rhythm 3-3-4-2-4. If a pulse was sounded it was replaced by a silent pulse, and vice-versa. The 17th rhythm (top) was the clave son itself. The mutations of the clave son rhythm are identified in Table 3 by the letters D and I denoting, respectively, a deletion and insertion of a sounded pulse.

The fourth data set consists of 13 rhythms composed of five onsets and 16 pulses each, shown in box notation in Table 4. All these rhythms are permutations of the adjacent IOIs of the popular clave son rhythm, with adjacent intervals 3-3-4-2-4. It should be emphasized that these permutation operations are not rotations or mirror reflections of the rhythms, but rather swaps of adjacent IOIs within the rhythms. For example consider the first two rhythms in Table 4: 24343 and 24433. The second is obtained from the first by swapping the third and fourth intervals, so that 34 in the first becomes 43 in the second. All these rhythms consist of two intervals of length three, two intervals of length four, and one interval of length two. These rhythms consist of a random subset of all possible such permutations. Therefore they do not necessarily have the same IOIs, but in some cases they do. The first column of Table 4 identifies the rhythms by their adjacent IOI sequences. This collection of five

Table 3. The seventeen 16-pulse Son-mutation rhythms.

| Rhythm ID | Binary Box Notation             |
|-----------|---------------------------------|
| Son       | x - - x - - x - - - x - x - - - |
| D-1       | - - - x - - x - - - x - x - - - |
| I-2       | x x - x - - x - - - x - x - - - |
| I-3       | x - x x - - x - - - x - x - - - |
| D-4       | x - - - - x - - - x - x - - -   |
| I-5       | x - - x x - x - - - x - x - - - |
| I-6       | x - - x - x x - - - x - x - - - |
| D-7       | x - - x - - - - - x - x - - -   |
| I-8       | x - - x - - x x - - x - x - - - |
| I-9       | x - - x - - x - x - x - x - - - |
| I-10      | x - - x - - x - - x x - x - - - |
| D-11      | x - - x - - x - - - - - x - - - |
| I-12      | x - - x - - x - - - x x x - - - |
| D-13      | x - - x - - x - - - x - - - - - |
| I-14      | x - - x - - x - - - x - x x - - |
| I-15      | x - - x - - x - - - x - x - x - |
| I-16      | x - - x - - x - - - x - x - - x |

intervals features prominently in Afro-Cuban music, and several of its permutations are well-known rhythms frequently used in practice, such as the Rumba and Gahu in Table 1. These 13 rhythms were selected at random from among all the possible permutations. Note that several rhythms are rotations of each other, such as 2-4-4-3-3 and 3-3-2-4-4 as well as 4-3-3-2-4 and 4-4-3-3-2. An important consequence of this property is that rotations or indeed mirror image reflections (or both) of a rhythm have identical IOI histograms. In other words, the IOIs of a rhythm are invariant under such transformations. Therefore methods that extract information from the full-interval histograms, either based on features or earth-mover-type of transformations, may not perform well depending on the fraction of rhythms in the data set

Table 4. The thirteen 5-onset Son-IOI permutation rhythms.

| Rhythm ID | Binary Box Notation             |
|-----------|---------------------------------|
| 24343     | x - x - - - x - - x - - - x - - |
| 24433     | x - x - - - x - - - x - - x - - |
| 32434     | x - - x - x - - - x - - x - - - |
| 32443     | x - - x - x - - - x - - - x - - |
| 33244     | x - - x - - x - x - - - x - - - |
| Son-33424 | x - - x - - x - - - x - x - - - |
| 34342     | x - - x - - - x - - x - - - x - |
| 34432     | x - - x - - - x - - - x - - x - |
| 42343     | x - - - x - x - - x - - - x - - |
| 43234     | x - - - x - - x - x - - - x - - |
| 43243     | x - - - x - - x - x - - - x - - |
| 43324     | x - - - x - - x - - x - x - - - |
| 44332     | x - - - x - - - x - - x - - x - |

that have this property. Furthermore, as the example of Fig. 1 makes clear, even rhythms that are not rotations or mirror images (or both) of each other may still have identical IOI histograms, thus exacerbating the problems for histogram-based methods.

## 6. Results and Discussion

To compare the various similarity measures with each other and with human judgments, the dissimilarity between every pair of rhythms in the data was first calculated, obtaining a dissimilarity (distance) matrix. Then a statistical procedure called the Mantel test<sup>17</sup> was used to calculate the correlation coefficient between pairs of these distance matrices using a software package developed by Bonnet and Van de Peer.<sup>4</sup> The Mantel test is a non-parametric permutation test specifically designed for matrices in which the elements are not independent, as is the case with this data, which consist of similarity judgments and distance-measure calculations between all the pairs of rhythms in a data set. In addition to handling any statistical dependencies that may be present between the elements within each matrix, the Mantel test has the nice feature that it is nonparametric, i.e. it makes no assumptions about the underlying probability distributions governing those elements. The results of the Mantel tests for the data sets in Tables 1–4 are listed in Tables 5–8, respectively, where the statistically significant results (including borderline significant) are indicated in boldface type. In each table, the first column lists the various measures of similarity that were tried. The middle column gives the correlations with the judgments made by human subjects, and the column on the right lists the correlation coefficients with the edit distance. The first method (Statistical Features) uses the first 12 statistical features used in music information retrieval<sup>9</sup> described in Sec. 4. The second method (Stat. Features and nPVI) uses all 13 features described in Sec. 4. The third method applies the normalized nPVI by itself. The four subsequent methods in rows 4–7 of each table indicate four different versions of the Mallows distance. The first method (Corpus-Norm. FH-Mallows Dist.) indicates that the histograms were normalized using the LCM of the numbers of cells in the histograms of *all* the rhythms in that corpus (data set). Furthermore, the prefix FH (denoting Full Histogram) indicates that the histogram was calculated using the distances between *all pairs* of onsets in the rhythms. The fifth method (Pair-Norm. FH-Mallows Dist.), is different from the previous one in that when the distance between a pair of rhythms is computed, the LCM is calculated only from the pair in question. If all the rhythms in a corpus have the same number of onsets, as is the case with the rhythms in Tables 5 and 8, then the two methods will necessarily give the same results. In the sixth and seventh methods the prefixes AH are used instead of FH. The prefix AH (denoting Adjacent Histogram) indicates that the histogram was calculated using only the distances between the temporally adjacent onsets in the rhythms. Finally, the last row in each table lists the correlation of the edit distance with the human judgments. For all the comparisons made in this paper, the human

Table 5. Mantel test results for the six Afro-Cuban 5-onset 16-pulse rhythms.

| Approach                      | Human Judgment                                       | Edit Distance            |
|-------------------------------|--|--------------------------|
| Statistical Features          | $r = -0.070$ $p = 0.473$                             | $r = -0.144$ $p = 0.283$ |
| Stat. Features and nPVI       | $r = 0.024$ $p = 0.439$                              | $r = -0.093$ $p = 0.417$ |
| nPVI only                     | $r = 0.238$ $p = 0.205$                              | $r = 0.065$ $p = 0.433$  |
| Corpus-Norm. FH-Mallows Dist. | <b><math>r = 0.699</math> <math>p = 0.022</math></b> | $r = 0.349$ $p = 0.200$  |
| Pair-Norm. FH-Mallows Dist.   | <b><math>r = 0.699</math> <math>p = 0.022</math></b> | $r = 0.349$ $p = 0.200$  |
| Corpus-Norm. AH-Mallows Dist. | $r = 0.572$ $p = 0.150$                              | $r = 0.220$ $p = 0.450$  |
| Pair-Norm. AH-Mallows Dist.   | $r = 0.572$ $p = 0.150$                              | $r = 0.220$ $p = 0.450$  |
| Edit Distance                 | <b><math>r = 0.767</math> <math>p = 0.017</math></b> | —                        |

judgment scores were obtained from the listening tests performed in previous studies carried out at the Radcliffe Institute for Advanced Study as well as the Music Department at Harvard University.<sup>20,34</sup> The reader is referred to those papers for detailed descriptions of the subjects used, the stimulus properties, and how the human listening tests were carried out.

Of the tests performed with the six Afro-Cuban 5-onset 16-pulse rhythms (Table 5), only the Mallows distance gave statistically significant results, correlating highly with human judgments ( $r = 0.699$ ,  $p = 0.022$ ). This is almost as high as the result obtained previously with the edit distance ( $r = 0.767$ ,  $p = 0.017$ ) calculated directly on the rhythms themselves.<sup>34</sup> Note that since in this corpus all the rhythms have the same number of onsets, corpus normalization is equivalent to pairwise normalization, and the two methods necessarily yield the same results.

By contrast with the above data, Rey’s nine Afro-Cuban variable-onset 8-pulse rhythms (Table 6), yielded moderate and mild but statistically significant correlations with human judgments for all the methods with the exception of Pair-Norm. AH-Mallows Distance. The nPVI, previously shown to be successful as a measure of musical rhythm complexity,<sup>32</sup> gave the lowest statistically significant correlation ( $r = 0.258$ ,  $p = 0.043$ ) when used in isolation, and all the other models yielded correlation coefficients ranging between 0.298 and 0.467. This represents a considerable drop from the result obtained with the edit distance ( $r = 0.590$ ,  $p = 0.0002$ ) when calculated directly on the rhythm sequences. Nevertheless, the nPVI does appear to capture some similarity information used by human listeners, and combining it with

Table 6. Mantel test results for Rey’s nine Afro-Cuban variable-onset 8-pulse rhythms.

| Approach                      | Human Judgment  | Edit Distance  |
|-------------------------------|---|--|
| Statistical Features          | <b><math>r = 0.438</math> <math>p = 0.005</math></b>    | <b><math>r = 0.577</math> <math>p = 0.003</math></b> |
| Stat. Features and nPVI       | <b><math>r = 0.467</math> <math>p = 0.003</math></b>    | <b><math>r = 0.572</math> <math>p = 0.003</math></b> |
| nPVI only                     | <b><math>r = 0.258</math> <math>p = 0.040</math></b>    | $r = -0.059$ $p = 0.435$                             |
| Corpus-Norm. FH-Mallows Dist. | <b><math>r = 0.448</math> <math>p = 0.004</math></b>    | <b><math>r = 0.413</math> <math>p = 0.021</math></b> |
| Pair-Norm. FH-Mallows Dist.   | <b><math>r = 0.298</math> <math>p = 0.038</math></b>    | <b><math>r = 0.540</math> <math>p = 0.007</math></b> |
| Corpus-Norm. AH-Mallows Dist. | <b><math>r = 0.322</math> <math>p = 0.024</math></b>    | <b><math>r = 0.509</math> <math>p = 0.005</math></b> |
| Pair-Norm. AH-Mallows Dist.   | $r = 0.193$ $p = 0.115$                                 | $r = 0.259$ $p = 0.077$                              |
| Edit Distance                 | <b><math>r = 0.594</math> <math>p &lt; 0.001</math></b> | —  |

the statistical features it bumps up the correlation from 0.438 to 0.467. However, there is no *a priori* reason to believe that in general, two rhythms should be perceived as similar if they have similar complexity values. Indeed, the nPVI does not correlate at all with the edit distance for any of the four data sets. Nevertheless the low but significant correlation between nPVI and human similarity judgments is noteworthy, and warrants further exploration of the relationship between these two parameters of musical rhythm. In this corpus the number of onsets in the rhythms varies considerably (between 2 and 6), and therefore the results with the corpus and the pairwise normalizations of the Mallows distance differ: using FH yields  $r = 0.448$ ,  $p = 0.004$  for corpus normalized, and  $r = 0.298$ ,  $p = 0.038$  for pairwise normalized, whereas using AH yields  $r = 0.322$ ,  $p = 0.024$  for corpus normalized, and  $r = 0.193$ ,  $p = 0.115$  for pairwise normalized. For this data set the statistical features calculated from the IOI histograms correlate quite highly with the edit distance ( $r = 0.577$ ,  $p = 0.004$ ).

The results with the statistical features are so different for the two data sets (compare Tables 5 and 6) that they warranted a deeper look at the feature values. Indeed, the range of values that these features take is very different for the two data sets. The average values of all the features are shown in Fig. 5, normalized to the [0-1] interval. The columns left of each pair belong to the 16-pulse rhythms of Table 5, and the ones on the right to the 8-pulse rhythms of Table 6. The error bars indicate  $\pm$  one standard deviation. It is abundantly clear that the range of values in each group differ greatly. For the 16-pulse rhythms four features have zero standard deviation: mean, total energy, centroid, and skewness. The other features are not much better in this regard. It appears that the number of distinct values that the statistical features can take is too small to provide them with much discriminating power. This

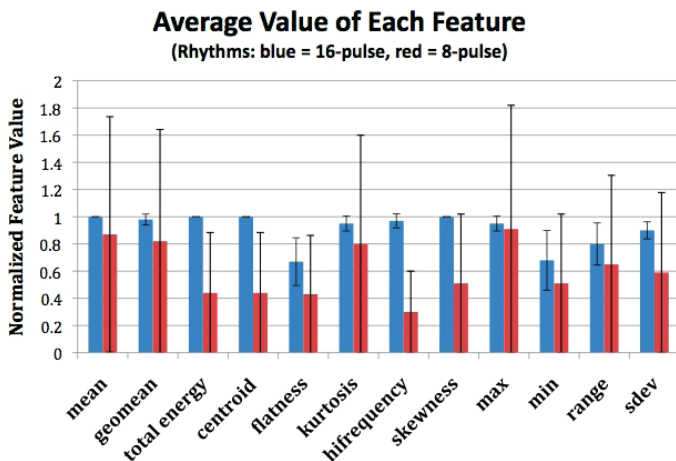


Fig. 5. The average value of each feature for the two data sets. Error bars indicate  $\pm$  one standard deviation (color online).

Table 7. Mantel test results for the seventeen 16-pulse mutation rhythms.

| Approach                      | Human Judgment          | Edit Distance            |
|-------------------------------|-------------------------|--------------------------|
| Statistical Features          | $r = 0.241$ $p = 0.026$ | $r = -0.022$ $p = 0.358$ |
| Stat. Features and nPVI       | $r = 0.263$ $p = 0.017$ | $r = -0.014$ $p = 0.358$ |
| nPVI only                     | $r = 0.220$ $p = 0.045$ | $r = 0.068$ $p = 0.521$  |
| Corpus-Norm. FH-Mallows Dist. | $r = 0.249$ $p = 0.048$ | $r = 0.045$ $p = 0.640$  |
| Pair-Norm. FH-Mallows Dist.   | $r = 0.249$ $p = 0.048$ | $r = 0.045$ $p = 0.640$  |
| Corpus-Norm. AH-Mallows Dist. | $r = 0.367$ $p = 0.002$ | $r = 0.083$ $p = 0.524$  |
| Pair-Norm. AH-Mallows Dist.   | $r = 0.367$ $p = 0.002$ | $r = 0.083$ $p = 0.524$  |
| Edit Distance                 | $r = 0.461$ $p = 0.055$ | —                        |

is likely the reason why they also fail to predict perception of similarity. The fact that this corpus is rather small is also likely to be a contributing factor.

Turning to the third data set (Table 7), all the methods are mildly but statistically significantly correlated with human judgments, and none correlate with the edit distance. Again, the nPVI appears to capture some similarity information used by human listeners ( $r = 0.220$  with  $p = 0.045$ ), and combining it with the statistical features it bumps up the correlation from  $r = 0.241$  with  $p = 0.026$  to  $r = 0.263$  with  $p = 0.017$ . The highest correlation is with the edit distance:  $r = 0.461$  with  $p = 0.055$ .

Table 8 shows the correlations obtained with the fourth data set consisting of the thirteen 5-onset, 16-pulse permutation rhythms. With this data set only the edit distance correlates with human judgments ( $r = 0.430$  with  $p = 0.001$ ). The nPVI and none of the methods based on IOI histograms correlate with either human judgments or the edit distance. The reason for this poor behavior will become clear following the discussion below.

In order to gain a better understanding of the general behavior of the performance of the IOI histogram-based methods with all four data sets, the corpus-normalized full-histogram Mallows distance matrices of the data sets were closely examined. A powerful visualization method for this purpose utilizes phylogenetic trees computed from the distance matrices. For each data set the *BioNJ* tree was computed using the *SplitsTree-4* package<sup>13</sup> and the resulting four trees are shown in Figs. 6–9. The nodes (leaves) in these trees represent the rhythms, and the length of the shortest path

Table 8. Mantel test results for the thirteen 16-pulse permutation rhythms.

| Approach                      | Human Judgment                | Edit Distance                 |
|-------------------------------|-------------------------------|-------------------------------|
| Statistical Features          | $r = -0.110$ $p = 0.175$      | $r = -0.109$ $p = 0.178$      |
| Stat. Features and nPVI       | $r = -0.069$ $p = 0.303$      | $r = -0.059$ $p = 0.318$      |
| nPVI only                     | $r = 0.014$ $p = 0.405$       | $r = 0.034$ $p = 0.352$       |
| Corpus-Norm. FH-Mallows Dist. | $r = -0.121$ $p = 0.137$      | $r = -0.152$ $p = 0.098$      |
| Pair-Norm. FH-Mallows Dist.   | $r = -0.121$ $p = 0.137$      | $r = -0.152$ $p = 0.098$      |
| Corpus-Norm. AH-Mallows Dist. | $r = 0.0001$ $p = \text{NaN}$ | $r = 0.0001$ $p = \text{NaN}$ |
| Pair-Norm. AH-Mallows Dist.   | $r = 0.0001$ $p = \text{NaN}$ | $r = 0.0001$ $p = \text{NaN}$ |
| Edit Distance                 | $r = 0.430$ $p < 0.001$       | —                             |

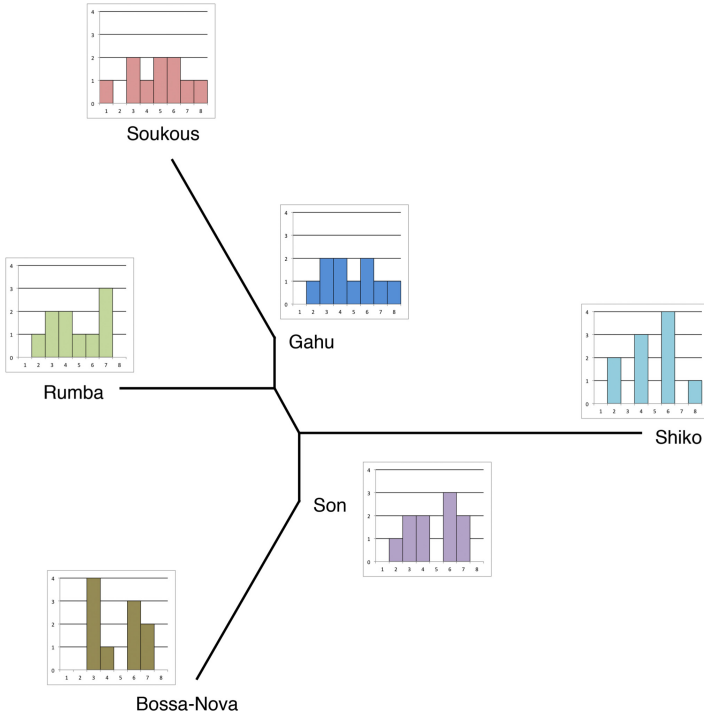


Fig. 6. *BioNJ* tree and IOI-histograms for the six Afro-Cuban rhythms with the corpus-normalized full-histogram Mallows distance.

(geodesic) along the tree branches between any two nodes is proportional to the entry in the distance matrix corresponding to the two rhythms associated with those nodes. Each node is labeled with the name of the rhythm assigned to that node, and a graph of the rhythm’s full-interval histogram is displayed next to each label. The *BioNJ* trees for both the edit distance and the human judgments for the six Afro-Cuban 5-onset, 16-pulse rhythms in Table 1, and the nine Afro-Cuban variable-onset 8-pulse rhythms in Table 2 may be found in the paper by Toussaint *et al.*<sup>34</sup> The *BioNJ* trees for both the edit distance and the human judgments for the seventeen 16-pulse Son-mutation rhythms in Table 3, and the thirteen 5-onset Son-IOI permutation rhythms in Table 4, may be found in the paper by Post and Toussaint.<sup>20</sup> These trees are not duplicated here for the sake of brevity, and the interested reader is referred to these papers for comparisons between those trees and the trees obtained with the Mallows distance given here. Furthermore, the *BioNJ* trees for the histograms of the statistical features listed in Eqs. (3)–(15) are not included here because the correlations with human judgments for this method were low, and therefore the trees would be quite distorted and unreliable.

From Fig. 6 it can be seen that all six Afro-Cuban 5-onset 16-pulse rhythms have distinct IOI histograms, thus giving histogram-based methods, at least in theory, a



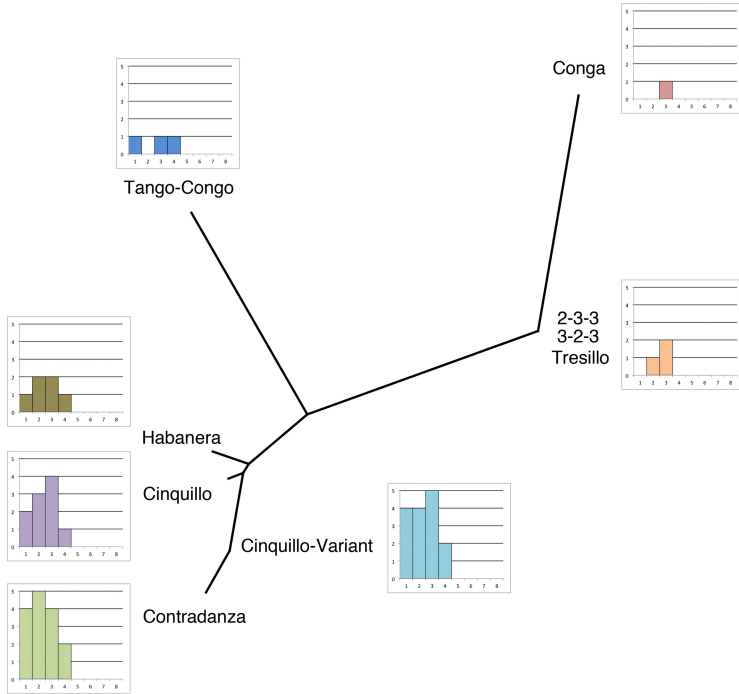


Fig. 7. *BioNJ* tree and IOI-histograms for Rey's nine Afro-Cuban rhythms with the corpus-normalized full-histogram Mallows distance.

chance to perform well. Indeed, the *BioNJ* tree exhibits structure that agrees with musicological knowledge: the son, gahu, and rumba are more similar to each other than the shiko, bossa-nova, and soukous. This is borne out by the human listening tests ( $r = 0.699$  with  $p = 0.022$ ), as well as by their proximity to each other in the tree. That the statistical features of the histograms fail to correlate with human judgments is probably due to the small size of the data set.

Concerning Rey's nine Afro-Cuban variable-onset 8-pulse rhythms, it may be observed from Fig. 7 that apart from the three 3-onset rhythms (tresillo, 3-2-3, and 2-3-3), the six other rhythms all have distinct histograms, thus again providing a reasonable performance test for histogram-based methods. This tree also provides a moderate agreement with musicological knowledge: the tresillo, 2-3-2, and 2-3-3 rhythms are very similar to each other, as are the cinquillo and its variant; in addition the tango-congo is more similar to the habanera than to the conga. These relationships are borne out by the relative distances between the nodes corresponding to these rhythms in the tree, and by the correlations with the matrices obtained from the listening tests, listed in Table 6.

Turning to the 13, 16-pulse mutation rhythms, it may be observed from Fig. 8 that the 17 rhythms yield only 10 distinct histograms. In this data set the son has five onsets, the 11 rhythms with an insertion (labeled with an I) have six onsets each, and

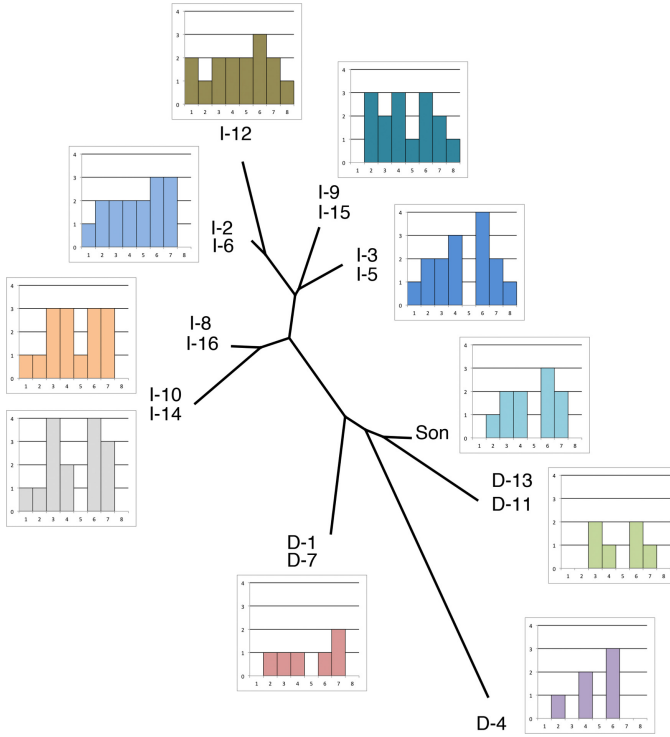


Fig. 8. *BioNJ* tree and IOI-histograms for the 17 mutation rhythms with the corpus-normalized FH Mallows distance.

the five rhythms with a deletion (labeled with a D) have four onsets each. Note also that only the son, D-4 and I-12 do not share the same histogram with any other rhythm. All the other rhythms have a partner with which they share the identical histogram: D-1 with D-7, D-11 with D-13, I-3 with I-5, I-9 with I-15, I-2 with I-6, I-8 with I-16, and I-10 with I-14. Therefore one would expect that for this data set the correlations with human judgments would be significantly lower than with the second data set. The correlations in Table 7 bear this out. Although all the correlations are statistically significant, they are mild correlations varying from 0.220 to 0.367. Considering the number of shared histograms among the rhythms this is a remarkable performance. The tree also exhibits several structural properties worth noting. First, the tree is composed of two large well-separated clusters: the insertion rhythms form one cluster at the top distinct from the cluster comprised of the deletion rhythms at the bottom. In addition, since these rhythms are all mutations of the son rhythm one would expect the son to be located in the central part of the tree; this is also evident in Fig. 8. The tree also has several commonalities with the tree obtained from the human judgments in a previous study.<sup>20</sup> Several groups of rhythms fall in the same clusters in both trees: I-2, I-3, and I-5, I-14 and I-16, and D-1, D-4 and D-7. Furthermore, several pairs of rhythms clustered together in the

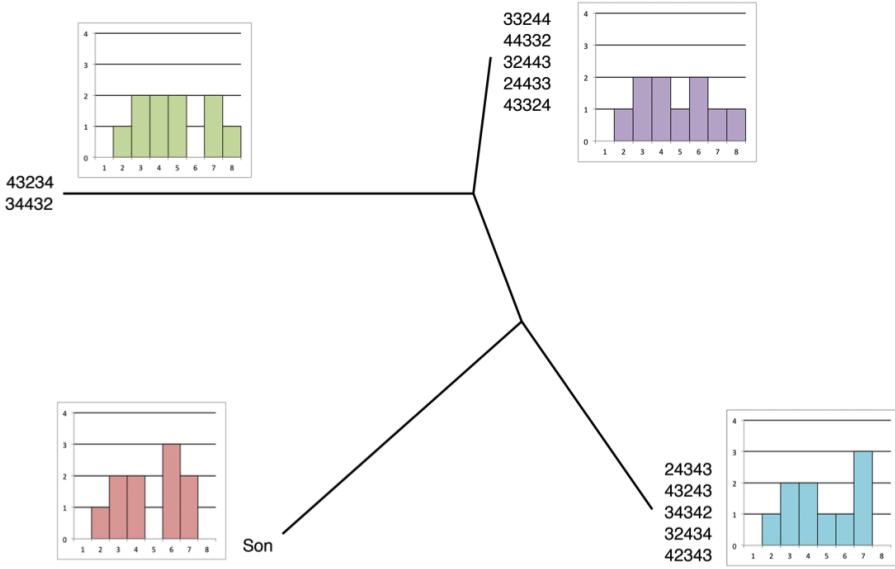


Fig. 9. *BioNJ* tree and IOI-histograms for the 13 permutation rhythms with the corpus-normalized full-histogram Mallows distance.

human judgment tree<sup>20</sup> have identical histograms, namely, I-15 and I-9, D-11 and D-13 and I-3 and I-5.

The *BioNJ* tree of the thirteen 16-pulse permutation rhythms shown in Fig. 9 yields only four different IOI histograms, and hence only four nodes in the tree, rather than 13. There are two large clusters of five rhythms each. The five rhythms in the upper right contain adjacent intervals of length three (33) and four (44). The five rhythms in the lower right contain no adjacent intervals of any length. The cluster on the upper left contains two rhythms with only adjacent intervals of length four (44). Finally, on the lower left the clave son (33424) is off by itself with only adjacent intervals of length three (33). At a higher level the tree exhibits two clusters. The clave son forms one larger cluster with the rhythms that contain no adjacent intervals of any length (lower right), and the two clusters at the top form the other large cluster. Interestingly these two large clusters may also be distinguished by the maximum height of the histograms: 2 for the top cluster and 3 for the bottom cluster. Since the 13 different rhythms map to only four different IOI-histograms any method that extracts information from only the IOI-histograms will necessarily perform poorly, and Table 8 bears this out. The results with this data set also highlight the robustness of the edit distance. In spite of the fact that the rhythms sound so similar, and that there are only four distinct histograms, the edit distance nevertheless correlates moderately and statistically highly significantly with human judgments ( $r = 0.430$ ,  $p < 0.001$ ).

## 7. Conclusions

One of the main conclusions drawn from this study is that the statistical features calculated from the IOI-histograms, as used by Gouyon *et al.* in the context of music information retrieval in the *acoustic* domain,<sup>9</sup> are, in the *symbolic* domain, much better than the music-theoretical structural features<sup>33</sup> for predicting human judgments of rhythm similarity for the nine variable-onset 8-pulse Afro-Cuban rhythms listed in Table 2. Nevertheless, they are still not as good as the edit distance in this regard. The results obtained here provide additional evidence of the superiority of the edit distance over other methods for measuring symbolically notated musical rhythm similarity. The edit distance performed well on all four data sets, even for the permutation rhythms that tend to sound quite similar to each other, and make the comparative listening tests difficult for human subjects.

Previous results have suggested that the structural features of musical rhythm that inform much discourse in music theory are incapable of predicting human judgments of rhythm similarity.<sup>33</sup> The poor results obtained here with the statistical features of the IOI histograms add further weight to the hypothesis that transformation methods are superior to feature-based methods as tools for predicting human judgments of similarity in general, and musical rhythm similarity in particular. Nevertheless, not much information appears to be lost when using the IOI histograms of the rhythms, rather than the rhythm sequences themselves, if the Mallows and generalized Mallows distances are used to compare the histograms rather than the statistical features, as long as the rhythms being compared have distinct IOI histograms. Indeed, the properties of the permutation data set highlight rather forcefully the fact that musical rhythms which are normally perceived as being different, may have identical IOI histograms. When this happens any method based on extracting information from only the IOI histograms is naturally doomed to failure.

The Mallows distances based on the IOI histograms correlate with the edit distance only for the third data set consisting of nine Afro-Cuban variable-onset 8-pulse rhythms. This suggests the speculative hypothesis that the corpus-normalized FH Mallows distance performs well for data sets that consist of rhythms that have a variable number of onsets. Further experiments with other data sets that have this property should be performed to test further this hypothesis.

The fact that human subjects considered as similar some of the pairs of rhythms that share the same IOI histograms suggests the additional hypothesis that human subjects are able to recognize rhythms with similar histograms in certain situations. It would be interesting to carry out experiments to test this hypothesis.

## Acknowledgments

This research was supported by a grant from the Provost's Office of New York University Abu Dhabi, through the Faculty of Science, in Abu Dhabi, The United Arab Emirates. Preliminary work was done in the fall of 2012 in Abu Dhabi, and the

research was completed while the first and last authors were visiting the Courant Institute of Mathematical Sciences at New York University in New York City.

## References

1. S. Arom, *African Polyphony and Polyrhythm* (Cambridge University Press, Cambridge, UK, 1991).
2. S. Arom, More on rhythmical marking: A reply to Herve Rivière, *Ethnomusicology* **38**(2) (1994) 321–322.
3. J. F. Beltran, X. Liu, N. Mohanchandra and G. T. Toussaint, Measuring musical rhythm similarity: Statistical features versus transformation methods, *Proc. Second Int. Conf. Pattern Recognition Applications and Methods* (Barcelona, Spain, 15–18 February 2013), pp. 595–598.
4. E. Bonnet and Y. Van de Peer, zt: A software tool for simple and partial Mantel tests, *J. Stat. Softw.* **7**(10) (2002) 1–12.
5. S.-H. Cha and S. N. Srihari, On measuring the distance between histograms, *Pattern Recogn.* **35** (2002) 1355–1370.
6. S. Dubuisson, Tree-structured image difference for fast histogram and distance between histograms computation, *Pattern Recogn. Lett.* **32** (2011) 411–422.
7. R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, 2nd edn. (Wiley-Interscience, 2000).
8. K. Fichtner, Homometric polytopes in cadmium iodide, *Acta Crystallogr. A* **42** (1986) 98–101.
9. F. Gouyon, S. Dixon, E. Pampalk and G. Widmer, Evaluating rhythmic descriptors for musical genre classification, *Proc. 25th International AES Conf.* (Audio Engineering Society, London, 2004), pp. 196–204.
10. E. Grabe and E. Low, Durational variability in speech and the rhythm class hypothesis, in *Papers in Laboratory Phonology 7*, eds. C. Gussenhoven and N. Warner (Cambridge University Press, Cambridge, 2002).
11. U. Hahn, N. Chater and L. B. Richardson, Similarity as transformation, *Cognition* **87** (2003) 1–32.
12. C. J. Hodgetts, U. Hahn and N. Chater, Transformation and alignment in similarity, *Cognition* **113** (2009) 62–79.
13. D. H. Huson, SplitsTree: A program for analyzing and visualizing evolutionary data, *Bioinformatics* **14**(10) (1998) 68–73.
14. E. Levina and P. Bickel, The earth mover’s distance is the Mallows distance: Some insights from statistics, in *Proc. Eighth IEEE Int. Conf. Computer Vision*, Vancouver, Canada, Vol. 2 (2001), pp. 251–256.
15. R. Lowrance and R. A. Wagner, An extension of the string-to-string correction problem, *J. Assoc. Comput. Mach.* **22** (1975) 177–183.
16. Y. Ma, X. Gu and Y. Wang, Histogram similarity measure using variable bin size distance, *Comput. Vis. Image Understand.* **114** (2010) 981–989.
17. N. Mantel and R. S. Valand, A technique of nonparametric multivariate analysis, *Biometrics* **26** (1970) 547–558.
18. N. J. Nilsson, *The Mathematical Foundations of Learning Machines* (Morgan Kaufmann Publishers, San Mateo, California, 1990).
19. A. L. Paterson, Ambiguities in the X-ray analysis of crystal structures, *Phys. Rev.* **64**(5–6) (1944) 195–201.

20. O. Post and G. T. Toussaint, The edit distance as a measure of perceived rhythmic similarity, *Empir. Musical. Rev.* **6**(3) (2011) 164–179.
  21. H. Rivière, On rhythmical marking in music, *Ethnomusicology* **37**(2) (1993) 243–250.
  22. J. Roeder, A calculus of accent, *Journal of Music Theory* **39**(1) (1995) 1–46.
  23. D. Sankoff and J. Kruskal (eds.), *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison* (CLSI Publications, Stanford University, 1999).
  24. E. D. Scheirer, R. B. Watson and B. L. Vercoe, On the perceived complexity of short musical segments, *Proc. Int. Conf. Music Perception and Cognition*, eds. C. Woods, G. Luck, R. Brochard, F. Seddon and J. Sloboda (Psychology Department, Keele University, UK, 2000).
  25. M. Senechal, A point set puzzle revisited, *Eur. J. Combinator.* **29**(1) (2008) 191–212.
  26. J. Serrà, T. H. Özasan and J. L. Arcos, Note onset deviations as musical piece signatures, *PLoS ONE* **8**(7) (2013) e69268, doi: 10.1371/journal.pone.0069268.
  27. F. Serratosa and A. Sanfeliu, Signatures versus histograms: Definitions, distances and algorithms, *Pattern Recogn.* **39** (2006) 921–934.
  28. R. N. Shepard, Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space, *Psychometrika* **22** (1957) 325–345.
  29. S. A. Sloman and L. J. Rips, *Similarity and Symbols in Human Thinking* (MIT Press, 1998).
  30. V. M. Sloutsky, The role of similarity in the development of categorization, *Trends Cogn. Sci.* **7**(6) 246–251.
  31. G. T. Toussaint, *The Geometry of Musical Rhythm* (Chapman-Hall/CRC Press, 2013).
  32. G. T. Toussaint, The pairwise variability index as a tool in musical rhythm analysis, *Proc. 12th Int. Conf. Music Perception and Cognition (ICMPC)*, and *8th Triennial Conf. European Society for the Cognitive Sciences of Music (ESCOM)* (Thessaloniki, Greece, 23–28 July 2012), pp. 1001–1008.
  33. G. T. Toussaint, L. Matthews, M. Campbell and N. Brown, Measuring musical rhythm similarity: Transformation versus feature-based methods, *J. Interdiscipl. Music Studies* **6**(1) (2012) 23–53.
  34. G. T. Toussaint, M. Campbell and N. Brown, Computational models of symbolic rhythm similarity: Correlation with human judgments, *Anal. Approaches World Music* **1**(2) (2011) 381–430.
  35. A. Tversky, Features of similarity, *Psychol. Rev.* **84** (1977) 327–352.
  36. R. R. Typke, R. C. Velkamp and F. Wiering, Searching notated polyphonic music using transportation distances, *Proc. ACM Multimedia Conf.* (New-York, USA, 2004), pp. 128–135.
-



**Juan Felipe Beltran** is part of the first generation of NYU Abu Dhabi alumni. Currently working as a research assistant for NYUAD's Design Technology Lab, his research interests include human-computer interaction, cognition, and artificial intelligence. He is

the recipient of the Courant Institute's Max Goldstein Award for Creativity in Computer Science, an avid debater, and an enthusiastic python teacher.



**Xiaohua Liu** was born and raised in Shanghai. Xiaohua graduated as part of the first cohort of NYU Abu Dhabi. He also spent more than a year studying and doing research at NYU's campus in New York. He worked as a research assistant in the Applied Mathematics

Laboratory and in the Center for Atmosphere Ocean Science at NYU. Xiaohua has presented at international conferences on neural information processing as well as pattern recognition. In his free time Xiaohua enjoys sports and traveling.



**Nishant Mohanchandra** received his B.S. degree in Computer Science from NYU Abu Dhabi and is part of its inaugural graduating class. At NYU Abu Dhabi, he was deeply involved in student affairs, serving as a resident adviser as well as an active student interest

group leader. He enjoys coding, teaching, creative writing, graphic novels, video games, sci-fi literature, and baking.



**Godfried Toussaint** is a Professor and the Head of the Computer Science Program at New York University Abu Dhabi in the United Arab Emirates, as well as an affiliate researcher in the Computer Science and Artificial Intelligence Laboratory at the Massa-

chusetts Institute of Technology in Cambridge, MA, USA. After obtaining his Ph.D. from the University of British Columbia in Vancouver, Canada, he taught and did research in the School of Computer Science at McGill University in Montreal, in the areas of information theory, pattern recognition, textile-pattern analysis and design, computational geometry, machine learning, music information retrieval, and computational music theory. In 2005, he became a researcher in the Centre for Interdisciplinary Research in Music Media and Technology, in the Schulich School of Music at McGill University. He is a founder and co-founder of several annual international conferences and workshops, including the ACM Symposium on Computational Geometry, and the Canadian Conference on Computational Geometry. He is an editor of several journals, including Computational Geometry: Theory and Applications, the International Journal of Computational Geometry and Applications, ISRN Geometry, and the Journal of Mathematics and the Arts. He received several distinguished awards including a Killam Fellowship from the Canada Council for the Arts, and a Radcliffe Fellowship from Harvard University in 2009, where he spent one year at the Radcliffe Institute for Advanced Study, and one year in the Music Department. His research on the phylogenetic analysis of musical rhythms has been reported in several media, and was the focus of two Canadian television programs. His latest book is titled: *The Geometry of Musical Rhythm*, published by Chapman-Hall/CRC Press, January 2013.