Original Paper _____ *Forma*, **21**, 129–149, 2006

A Comparison of Rhythmic Dissimilarity Measures

Godfried TOUSSAINT

School of Computer Science, McGill University, Montréal, Québec, Canada
E-mail address: godfried@cs.mcgill.ca

(Received February 25, 2005; Accepted June 10, 2006)

Keywords: Binary Sequences, Musical Rhythm, Rhythm Dissimilarity Measures, Music Information Retrieval, Phylogenetic Analysis

Abstract. Measuring the dissimilarity between musical rhythms is a fundamental problem with many applications ranging from music information retrieval and copyright infringement resolution to computational music theory and evolutionary studies of music. A common way to represent a rhythm is as a binary sequence where a zero denotes a rest (silence) and a one represents a beat or note onset. This paper first reviews methods for visualizing and representing rhythms, and then compares various measures of rhythm dissimilarity, including the Hamming distance, the Euclidean interval-vector distance, the interval-difference distance measure of Coyle and Shmulevich, the swap distance, and the chronotonic distance measures of Gustafson and Hofmann-Engl. Traditionally, rhythmic dissimilarity measures are compared according to how well rhythms may be recognized with them, how efficiently they can be retrieved from a data base, or how well they model human perception and cognition of rhythms. In contrast, here dissimilarity measures are compared on the basis of how much insight they provide about the structural inter-relationships that exist within families of rhythms, when phylogenetic trees and graphs are computed from the distance matrices determined by these dissimilarity measures. For two fundamental collections of rhythms, namely the 4/4 time and 12/8 time clave-bell time lines used in traditional African and Afro-American music, the chronotonic and swap distances appear to be superior to the other measures, and each has its own attractive features.

1. Introduction

Imagine that you are at a concert, after a stunning performance, clapping at a fast uniform pace, much like a heart beat while jogging. Even better, try it out right now, but stop after you reach sixteen claps. Then do it again but this time execute the first, fourth, seventh, eleventh and thirteenth claps loudly, and the remaining eleven claps softly. Your clapping pattern could then be represented like this: [**1** 2 3 **4** 5 6 **7** 8 9 10 **11** 12 **13** 14 15 16], where the claps shown in bold face are the ones you clap loudly. It may also be helpful to count aloud the sixteen claps in groups as follows: [**1** 2 3 **1** 2 3 **1** 2 3 4 **1** 2 **1** 2 3 4], and to clap loudly only on the one's. If you repeat this pattern over and over, making the soft claps completely inaudible, you will be clapping one of the most popular rhythms on the

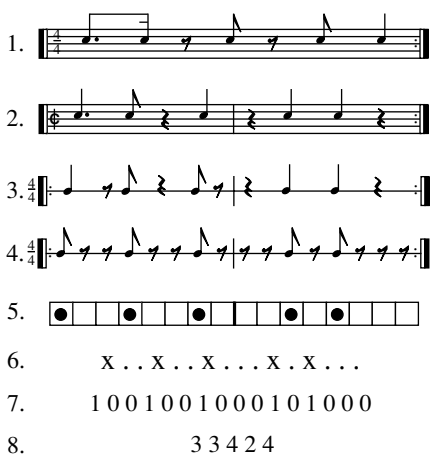

Fig. 1. Eight common ways of representing the clave *Son* rhythm.

planet, known mainly as the *Clave Son* from Cuba (TOUSSAINT, 2002).

The *Clave Son* rhythm is usually notated for musicians using standard music notation which affords many ways of expressing a rhythm. Four such examples are given in the top four lines of Fig. 1. The fourth line displays the rhythm with music notation using the smallest convenient durations of notes and rests. Western music notation is not ideally suited to represent African music (EKWUEME, 1974; AROM, 1991). The fifth and sixth lines show two popular ways of representing rhythms that avoid Western notation. The representation on line five is called the *Box Notation Method* made popular by Philip Harland at the University of California in Los Angeles in 1962, and is also known as TUBS (Time Unit Box System). The TUBS representation is prefereed among ethnomusicologists, and is invaluable to percussionists not familiar with Western notation (EKWUEME, 1974). It is also convenient for experiments in the psychology of rhythm perception, where a common variant of this method is simply to use one symbol for the beat and another for the pause, as illustrated in line six. In physiology, where the study of cardiac rhythms is important, as well as in computer science the clave *Son* would be written as the 16-bit binary sequence shown on line seven. Finally, line eight depicts the interval length representation of the clave *Son*, where the numbers denote the lengths (in shortest convenient units) of the intervals between consecutive beats (onsets). The compactness and ease of use in text, of this numerical interval-length representation, are two of its obvious advantages, but its iconic value is minimal.

This paper first reviews methods for visualizing and representing rhythms, and then compares various measures of rhythm dissimilarity, including the Hamming distance, the Euclidean interval-vector distance, the interval-difference distance measure of Coyle and Shmulevich, the swap distance, and the chronotonic distance measures of Gustafson and Hofmann-Engl. The question of which measure to use in a particular application has received a lot of attention in all areas that are concerned with the comparison of sequences.

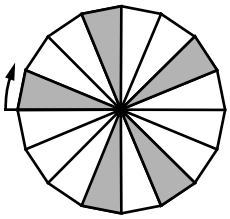

Fig. 2. The clave *Son* represented in 13th Century Arabic cyclic notation.

One rationale discussed in (KRUSKAL, 1983) is the application of the universal *post hoc* justification process. To quote Kruskal, "If we use any particular definition for distance, and find that this kind of distance supplies information we want, that "it works" when we check its performance, then the satisfactory performance justifies the definition." This rationale implies the comparison of several measures to determine which works best for the intended application. Traditionally, rhythmic dissimilarity measures have been compared according to how well rhythms may be recognized with them, how efficiently the measures retrieve query-rhythms from a data base, or how well they model human perception and cognition of rhythms. In contrast, here, dissimilarity measures are compared on the basis of how much insight they provide about the structural inter-relationships that exist within families of rhythms, when phylogenetic trees and graphs are computed from the distance matrices determined by these same dissimilarity measures. Experiments are reported for two fundamental collections of rhythms, namely the 4/4 time and 12/8 time clave-bell time lines used in traditional African and Afro-American music. These experiments lead to the conclusion that the chronotonic and swap distances appear to be superior to the other measures, and each has its own attractive features. Although these data sets are small in terms of the number of rhythms they contain, they were used because they are the most important rhythms in some sense, and we understand them well. This knowledge allows us to better judge the merits of the results.

2. Geometric Representation of Rhythms

Because traditional Western music notation is not well suited for the visual representation of rhythm, scholars of rhythm have sought to develop graphical geometrical representations that afford not only easier manipulation and visualization of rhythms, but that are applicable to domains other than music. However, geometric representations of rhythms have existed for quite some time even before Western music notation was well developed. For example, an Arabic book about rhythm written by Safi-al-Din in 1252 depicts rhythms using a circle divided into "pie slices" (WRIGHT, 1978). The *Son* represented in this notation is depicted in Fig. 2, where the pie slices that correspond to beats are shown shaded, and an arrow indicates the starting note and the direction of the flow of time.

Kjell Gustafson, at the Phonetics Laboratory of the University of Oxford, published

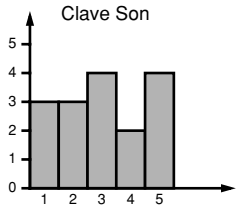

Fig. 3. The *adjacent-interval-spectrum* of the clave *Son*.

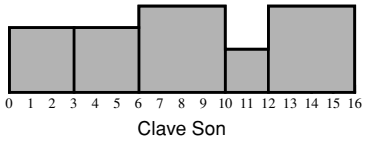

Fig. 4. The TEDAS representation of the clave *Son*.

an original method to represent a rhythm as a two-dimensional graph (GUSTAFSON, 1988). His idea (first proposed in 1983) is best explained with an example. Consider first the clave *Son* pattern shown in box notation on line five of Fig. 1. The interval lengths (3, 3, 4, 2, 4) between the onsets of the beats are ordered along the *x*-axis as a one-dimensional pattern. Although this representation is realistic in terms of the time at which beats occur, the relative durations of the intervals are not easily observed. In a *histogram* approach to rhythm visualization the intervals are plotted along the *y*-axis (GUSTAFSON, 1988), resulting in the *adjacent-interval-spectrum* of the rhythm. Figure 3 depicts the clave *Son* with such a representation. Here the relative lengths of the intervals are clearly visible but the temporal information along the *x*-axis is lost. To obtain a graphical representation that posesses the advantages of both of these methods, Gustafson simply combines them. The result of this union is illustrated with the clave *Son* in Fig. 4. Each temporal element (interval) is now a box and both the *x* and *y* axes represent the length of time of the interval. Gustafson refers to such a display as TEDAS (Temporal Elements Displayed As Squares). A TEDAS representation of rhythm has other advantages besides those mentioned in the preceeding. Some of these will be discuss2mann-Engl (HOFMANN-ENGL, 2002) rediscovered the idea of representing rhythms as a two-dimensional display where both axes measure time. Whereas Gustafson was motivated by visualizability and technical considerations, Hofmann-Engl, on the other hand, was concerned with predicting human perception and cognition of rhythmic dissimilarity. Althought the language used by Hofmann-Engl is quite different from that employed by Gustafson, their representations are equivalent. As in the preceeding, the idea is illustrated here with the clave *Son* as an example.

First the rhythm is broken down into its smallest time units (16th notes and rests in this case) as in line four of Fig. 1. Hofmann-Engl calls these atomic beats. Then each atomic beat is assigned a *y*-coordinate (height) equal to the length of the beat-to-beat interval in

Fig. 5. The clave *Son* as a *chronotonic chain.*

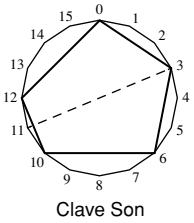

Fig. 6. The clave *Son* represented as a convex polygon.

which it falls, resulting in a pattern of sixteen points in the plane. Finally these points are connected together in increasing x-coordinate to obtain an x-monotonic polygonal chain. Hofmann-Engl calls such polygonal curves *chronotonic chains*. Figure 5 shows the clave *Son* represented as a chronotonic chain.

A third geometric method of representing a rhythm is illustrated in Fig. 6 where the clave *Son* is depicted as a convex polygon. Like the chronotonic representations of Gustafson and Hofmann-Engl, the polygon representation clearly shows the relative lengths of the onset or beat intervals, and is realistic on the circular time scale. In fact this representation is even more realistic than its chronotonic counterparts because rhythm timelines are cyclic, and all the other representations discussed above, apart from the Arabic "pie-slice" representation, have the beginning and end of the pattern located physically distant from each other, when they actually coincide cyclically in time. This representation also yields a powerful visualization tool, since a variety of geometric properties of rhythms are immediately perceived from the resulting polygons. For example, the dashed line connecting locations 3 and 11 in Fig. 6 indicates an axis of mirror symmetry for the clave *Son*. The polygon can also be used to compute a variety of n-dimensional vectors of geometric features $X = (x_1, x_2, ..., x_n)$. For example, focusing on the vertices of the *Son* polygon $V = (v_1, v_2, v_3, v_4, v_5)$ one could take v_i to be the internal angle of the i-th vertex. On the other hand viewing the polygon as a sequence of edges $E = (e_1, e_2, e_3, e_4, e_5)$ one could take e_i to be the length of the i-th edge of the polygon. In addition, one could use a variety of global shape features of the polygon such as the ratio of perimeter to area, or various moments of inertia. Recently, geometric features such as these have been used to obtain new classifications of rhythm timelines in African and Afro-American traditional music (TOUSSAINT, 2002, 2003a). Such classifications fall under the more recent *structural*

approach proposed by Jay Rahn (RAHN, 1987, 1996), and contrast sharply with earlier statistical methods (for example, Abraham Moles (MOLES, 1966), classified rhythms based on the entropy function of the histogram of relative frequencies of occurrences of their inter-onset intervals).

3. Measuring the Similarity of Rhythms

At the heart of any algorithm for comparing, recognizing or classifying rhythms, lies a measure of the dissimilarity between a pair of rhythms. The type of dissimilarity measure chosen is in part predetermined by the manner in which the rhythm is represented. Furthermore, the design of a measure of dissimilarity is guided by at least two fundamental ideas: *what* should be measured, and *how* should it be measured.

There exists a wide variety of methods for measuring the dissimilarity of two rhythms represented by strings of symbols (TOUSSAINT, 2003b). Indeed the resulting approximate pattern matching problem is a classical problem in pattern recognition and computer science in general. Traditionally dissimilarity between two patterns is measured by a simple template matching operation. More recently dissimilarity has been measured with more powerful and complex functions such as the *earth mover's distance* (CHA and SRIHARI, 2002; TYPKE *et al.*, 2003).

In this section we describe the five, easy to compute, popular and representative measures of rhythm dissimilarity that will be compared in the following.

3.1. The Hamming distance

For binary sequences a natural measure of distance or dissimilarity is the Hamming distance widely used in coding theory (HAMMING, 1986). The Hamming distance is simply the number of places in the strings where elements do not match. When a measure of similarity is preferred some researchers use the number of matches instead (MEUDIC, 2003). Hence it is one of the simplest kinds of template matching. Note that in this approach each rhythm is represented by a vector $X = (x_1, x_2, ..., x_n)$ where the x_i represent binary valued features of the rhythm. If a note is played at time unit i then $x_i = 1$ and otherwise $x_i = 0$. Thus rhythms are represented as points in an n-dimensional binary-valued vector space (usually called a *hypercube*). The Hamming distance between two points $X = (x_1, x_2, ..., x_n)$ and $Y = (y_1, y_2, ..., y_n)$ in this space is given by:

$$d_H(X,Y) = \sum_{i=1}^{n} |x_i - y_i| \tag{1}$$

where $|z|$ denotes the absolute value of z.

The Hamming distance is easily computed in $O(n)$ time. However, it is not very appropriate for our problem of rhythm dissimilarity, when used with this particular type of representation, because although it measures the existence of a missmatch, it does not measure how far the missmatch occurs. Furthermore, if a note or beat is moved a large distance the resulting modified rhythm will sound more different than if the beat is moved a small distance.

3.2. The Euclidean interval vector distance

Some rhythm detection algorithms and systems for machine recognition of music patterns use *inter-onset intervals* as the basic features for measuring dissimilarity (MONT-REYNAUD and GOLDSTEIN, 1985; COYLE and SHMULEVICH, 1998). These are the intervals of time between consecutive note onsets in a rhythm. This approach and its variants are more appropriate than the Hamming distance for measuring the dissimilarity of rhythms. Each rhythm may be represented by a vector of numbers that characterize these intervals. More specifically a rhythm may be represented by $X = (x_1, x_2, ..., x_n)$, where x_i is the number of vertices skipped by the i-th polygon edge starting at vertex labelled 0. This is essentially the same as the sequence of inter-onset time intervals since the time interval is the number of vertices skipped plus one. The dissimilarity between two rhythms $X = (x_1, x_2, ..., x_n)$ and $Y = (y_1, y_2, ..., y_n)$ may then be measured by the Euclidean distance (or some other metric) between the two interval vectors X and Y, in this n-dimensional interval vector space, given by:

$$d_E(X,Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \tag{2}$$

The interval vectors may be computed from the given binary sequences in $O(n)$ time, after which $d_E(X, Y)$ can be computed in $O(k)$ time, where k is the number of onsets, thus leading to $O(n)$ time overall.

3.3. The interval-difference vector distance

Coyle, Shmulevich and colleagues (SHMULEVICH *et al.*, 2001) represent a music pattern by what they call a *difference-of-rhythm vector*. If $T = (t_1, t_2, ..., t_n)$ is a vector of inter-onset time intervals for the notes of a rhythm then they define the difference-of-rhythm vector as $X = (x_1, x_2, ..., x_{n-1})$, where $x_i = t_{i+1} = t_i$. In the context of cyclic rhythms considered here $x_n = t_1 = t_n$, and the dimensionality of the difference-of-rhythm vector is thus n rather than $n - 1$. We shall use the shorter and more descriptive term *interval-difference* vector for the difference-of-rhythm vector. Let $X = (x_1, x_2, ..., x_n)$ and $Y = (y_1, y_2, ..., y_n)$ denote two such interval-difference vectors. They define the dissimilarity between X and Y by:

$$d_{ID}(X,Y) = \left(\sum_{i=1}^{n} \frac{\max(x_i - y_i)}{\min(x_i - y_i)} \right) - n. \tag{3}$$

Clearly this distance measure is also easily computed in $O(n)$ time.

3.4. The swap distance

To combat the inherent weakness of the Hamming distance, many variants and generalizations of the Hamming distance have been proposed over the years. One early generalization is the edit distance which allows for insertions and deletions of notes

(MONGEAU and SANKOFF, 1990; ORPEN and HURON, 1992). A noteworthy more recent generalization is the *fuzzy* Hamming distance which allows *shifting* of notes as well as insertions and deletions (BOOKSTEIN *et al*., 2002). Using *dynamic programming* these distances may be computed in $O(n^2)$ time.

This completely different approach to measuring the dissimilarity between two strings computes the amount of "work" required to transform one string into the other. Such an approach is common in bioinformatics where the two strings to be compared are chain polymers and the work is measured by the minimum number of basic operations required to transform one molecule into the other. The type of basic operation used varies and usually models some kind of mutation relevant to evolution. Here we will use a simplified version of the fuzzy Hamming distance that is much more efficient to compute.

The problem of comparing two binary strings of the same length with the same number of one's suggests an extremely simple edit operation called a *swap*. A swap is an interchange of a one and a zero that are adjacent to each other in the binary string. Interchanging the position of elements in strings of numbers is a fundamental operation in many sorting algorithms (BRUIJN, 1974). However, in the sorting literature a swap may interchange non-adjacent elements. When the elements are required to be adjacent, the swap is called a *mini-swap* or *primitive-swap* (BIEDL *et al*., 2004). Here we use the shorter term *swap* for the interchange of two adjacent elements. The swap distance between two rhythms is the *minimum* number of swaps required to convert one rhythm to the other. For example the rhythm [x . x . x x . x . x . x] can be converted to the rhythm [x . x x . x x . x . x .] by a minimum of four swaps, namely interchanging the third, fifth, sixth, and seventh beats with the corresponding rests preceeding them. The *swap* distance may be viewed as a simplified version of the *fuzzy* Hamming distance, where only the shift operation is used, and the cost of the shift is equal to its length (BOOKSTEIN *et al*., 2002). Such a measure of dissimilarity appears to be more appropriate than the Hamming distance between the binary vectors or the Euclidean distance between the interval vectors, in the context of rhythm dissimilarity (TOUSSAINT, 2002, 2003a). The swap distance may also be viewed as a special case of the more general *earth mover's distance* (also called *transportation distance*) used to measure melodic dissimilarity (TYPKE *et al*., 2003). Given two sets of points called supply points and demand points, each assigned a weight of material, the earth movers distance measures the minimum amount of work (weight times distance) required to transport material from the supply points to the demand points. No supply point can supply more weight than it has and no demand point receives more weight than it needs. Typke *et al*. solve this problem using linear programming, a relatively costly computational method. The swap distance is a one dimensional version of the earth mover's distance with all weights equal to one. Furthermore, in the case where both binary sequences have the same number of "one's" (or onsets), there is a one-to-one correspondence between the indices of the ordered onsets of the sequences. For example, consider again the two sequences X = [x . x . x x . x . x . x] and Y = [x . x x . x x . x . x .], each with seven onsets. The i-th onset of X must travel the necessary distance to reach the position of the i-th onset of Y. For i = 1 this distance is zero. For i = 3 the distance is one.

The swap distance may of course be computed by actually performing the swaps, but this is inefficient. If X has one's in the first $n/2$ positions and zero's elsewhere, and if Y has one's in the last $n/2$ positions and zero's elsewhere, then at least a quadratic number of

swaps would be required. On the other hand, if we compare the distances instead, a much more efficient algorithm results. First scan the binary sequence and store a vector of the x-coordinates at which the k onsets occur. For example, X and Y above yield the vectors $U = (u_1, u_2, ..., u_7) = (1, 3, 5, 6, 8, 10, 12)$ and $V = (v_1, v_2, ..., v_7) = (1, 3, 4, 6, 7, 9, 11)$, respectively. The difference between u_i and v_i is the number of swaps that must be performed to bring the two onsets into alignment. Therefore, in general, the swap distance between two onset-coordinate vectors U and V with k onsets is given by:

$$d_{SWAP}(U,V) = \sum_{i=1}^{k} |u_i - v_i|. \tag{4}$$

Computing U and V from X and Y is done trivially in $O(n)$ time with a simple scan. Therefore $O(n)$ time suffices to compute $d_{SWAP}(U, V)$, resulting in a large gain over using linear or dynamic programming.

3.5. The chronotonic distance

Hofmann-Engl considers the chronotonic chain as a vector of *atomic* units rather than as a piece-wise linear function. For example, the clave *Son* which has the interval set (3, 3, 4, 2, 4) and corresponding chronotonic chain shown in Fig. 5, yields the chronotonic 16-dimensional vector (3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 2, 2, 4, 4, 4, 4). To measure the dissimilarity between two rhythms Hofmann-Engl computes a weighted Euclidean distance between their corresponding chronotonic vectors. Hofmann-Engl reports that experiments with human subjects supports the hypothesis that this measure agrees with human perceptual and cognitive models of rhythmic dissimilarity (HOFMANN-ENGL, 2002). However, viewing the chronotonic representation as a function, such as the piece-wise linear function in Fig. 4, opens the door to a large family of possible distance functions with a long history in the fields of statistics and pattern recognition (TOUSSAINT, 1974a). Given two probability density functions $f_1(x)$ and $f_2(x)$, there are many measures of the distance (or separation) between them. One of the most well known is the *discrimination information*. The discrimination information between $f_1(x)$ and $f_2(x)$, also called the *Kullback-Liebler divergence* between the distributions (BERENZWEIG *et al.*, 2003), is given by:

$$KL = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} dx. \tag{5}$$

This measure is closely related to the *Kolmogorov variational* distance (TOUSSAINT, 1975) given by:

$$K = \int |f_1(x) - f_2(x)| dx. \tag{6}$$

For surveys of these and other information measures, their properties, and their interrelationships the reader is referred to (TOUSSAINT, 1974b; MATHAI and RATHIE, 1975). One can of course also compute the Kolmogorov variational distance K with

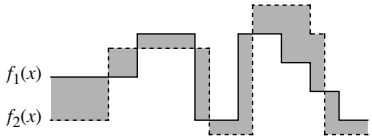

Fig. 7. Two melodies as rectilinear pitch-duration functions of time.

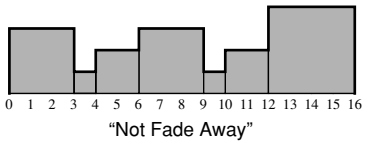

Fig. 8. The rhythm in *Not Fade Away* in chronotonic notation.

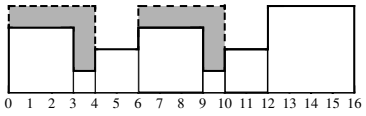

Fig. 9. The area difference K between *Not Fade Away* and *Shiko*.

functions that are not necessarily probability density functions. In fact, Ómaidín proposed exactly such a measure of dissimilarity between two melodies modelled as monotonic pitch-duration functions of time, as depicted in Fig. 7, where the x-axis measures time duration and the y-axis measures the pitch of the notes (ÓMAIDÍN, 1998). In the context of information retrieval of melodies one may also be interested in computing the minimum area between the two curves under translations of one curve in the x and y directions while the second curve remains fixed. Aloupis *et al*. show that for two melodies (polygonal chains) consisting of n vertices this function can be computed in $O(n^2 \log n)$ time (ALOUPIS *et al*., 2003).

Here we will use the measure K to compare rhythms using the chronotonic representation proposed by GUSTAFSON (1988). As an example, consider the rhythm [x . . x x . x . . x x . x . . .] shown in Fig. 8 in chronotonic notation. This is the rhythm timeline played on the *tom-tom* drum of *Not Fade Away* by Buddy Holly and the Crickets. The record was released in 1957 and became an instant hit (GOLDROSEN, 1975). It is a variant of the famous rhythm dubbed the *Bo Diddley beat* popularised by Bo Diddley in his 1955 hit "Bo Diddley," given by [x . x x x . x . . . x . x . . .], and also characterized by the mnemonics "*shave and a haircut, two bits*" as well as "*bump bubby bump bump, bump bump*". If we superimpose the *Not Fade Away* and *Shiko* timeline rhythms, as in Fig. 9, then the measure K, or area-difference (shown shaded) between the two curves, is equal to 12. A similar calculation superimposing

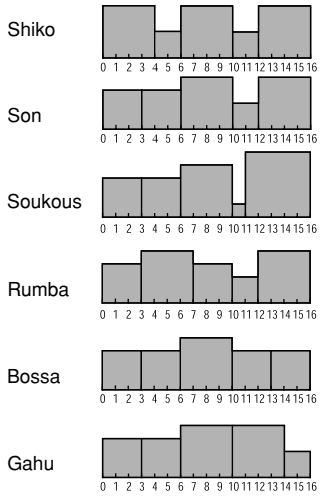

Fig. 10. The chronotonic representation of the 4/4 time *clave* patterns.

the *Not Fade Away* and clave *Son* timelines yields a value of K equal to 10, thus showing that the *Not Fade Away* rhythm is closer to the *Son* than the *Shiko*. In this discrete setting it is clear that K may be computed easily in $O(n)$ time.

4. Comparing Rhythmic Similarity Measures

Traditionally, rhythmic dissimilarity measures have been evaluated or compared with respect to how well rhythms may be recognized (COYLE and SHMULEVICH, 1998), how efficiently they can be retrieved from a data-base (FOOTE *et al.*, 2002), or how well they model human perception and cognition of rhythms (HOFMANN-ENGL, 2002; BERENZWEIG, 2003). In contrast, here we compare rhythmic dissimilarity measures with respect to how much insight they provide concerning the structural interelationships that exist within families of rhythms, when phylogenetic trees and graphs are computed from the distance matrices determined by these dissimilarity measures.

The dissimilarity measures discussed in the preceeding section are compared here using two families of rhythms, namely, the fundamental binary and ternary timelines used in African and AfroAmerican traditional music. In particular, the first family consists of the six five-beat clave-bell patterns in 4/4 time (TOUSSAINT, 2002), and the second family is comprised of the ten seven-beat bell patterns in 12/8 time used in West African traditional music (TOUSSAINT, 2003a). These two families are depicted in the chronotonic notation of Gustafson in Figs. 10 and 11, respectively.

4.1. Phylogenetic trees

Several techniques exist for generating phylogenetic trees from distance matrices (HUSON, 1998). Some of these methods have the desirable property that they produce

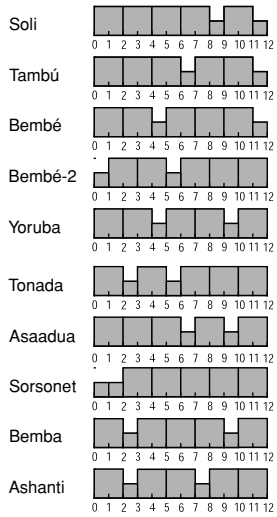

Fig. 11. The chronotonic representation of the 12/8 time *bell* patterns.

graphs that are not trees, when the underlying proximity structure is inherently not tree-like. One notable example is *SplitsTree* (DRESS *et al*., 1996). Like the more traditional phylogenetic trees, *SplitsTree* computes a plane graph embedding with the property that the distance in the drawing between any two nodes reflects as closely as possible the true distance between the corresponding two rhythms in the distance matrix. However, if the tree structure does not match the data perfectly then new nodes in the graph may be introduced, as for example in Fig. 12, with the goal of obtaining a better fit. Such nodes may suggest implied "ancestral" rhythms from which their "offspring" may be derived. In addition, edges may be split to form parallelograms, (or more general zonotopes) such as in Fig. 16. The relative sizes of these parallelograms are proportional to an *isolation index* that indicates how significant the clustering relationships inherent in the distance matrix are. *SplitsTree* also computes the *splitability index*, a measure of the goodness-of-fit of the entire splits graph. This fit is obtained by dividing the sum of all the approximate distances in the splits graph by the sum of all the original distances in the distance matrix.

5. Results

In this section we describe the structure of each *SplitsTree* obtained with each of the five distance measures described in the preceeding, when applied to the two families of rhythms. We begin with the binary (4/4 time) rhythms.

5.1. Binary rhythms

5.1.1 The Hamming distance

For the Hamming distance the splits graph shown in Fig. 12 is a tree with a perfect fit of 100%. The *Son* pattern lies at the center of the tree according to several definitions of

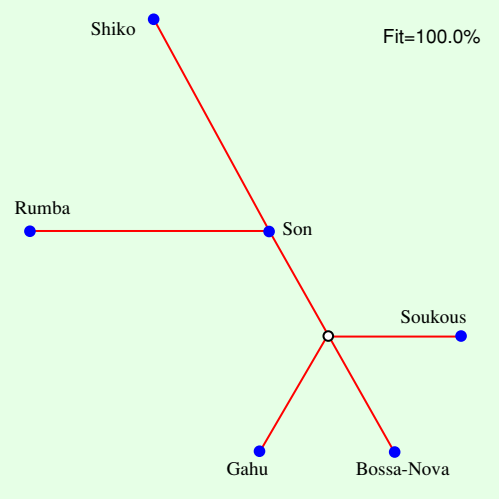

Fig. 12. The SplitsTree with the Hamming distance.

the notion of center. For example, the *Hamming Center* of a set *S* of binary strings is defined as a binary string that minimizes the maximum Hamming distance to all strings in *S* (GASIENIEC *et al*., 1999). If the search for a minimum is restricted to the set *S* then the resultant string is the *Set Hamming Center* of *S*. The reader may verify that the *Son* pattern is the *Set Hamming Center* of this family of six rhythms. As a second example, consider the *Hamming Median* of a set *S* of binary strings, defined as a binary string that minimizes the sum of the Hamming distances to all strings in *S* (JIANG *et al*., 2003). This measure is sometimes also called the *generalized Hamming* median. In addition, if the search for a minimum is restricted to the set *S* then the resultant string is the *Set Hamming Median* of S. The reader may verify that the *Son* pattern is also the *Set Hamming Median* of this family of six rhythms. The Hamming distance between the *Son* and each of the other rhythms is equal to *two*. The tree in Fig. 12 also contains one *ancestral* node with *Son*, *Soukous*, *Bossa-Nova*, and *Gahu* as its offspring, each with a Hamming distance of one from the ancestral rhythm. This ancestral rhythm is the four-beat pattern [x . . x . . x . . . x . . . . .]. Note that all four offspring contain this pattern, and they differ only in the location of the last (fifth) beat that falls in positions 12, 13, 14, and 15, for the Soukous, *Son*, *Bossa-Nova*, and *Gahu* rhythms, respectively. Finally, it is worth observing that there is no unique pair of candidates that represents the two most different rhythms.

5.1.2 The Euclidean interval vector distance

For the Euclidean interval vector distance the splits graph shown in Fig. 13 is tree-like with a good fit of 82.3%. The center of the graph is the *Son*, but it does not have direct paths to all other rhythms. The path from *Son* to *Gahu* is via the *Bossa-Nova*. Here there are two rhythms, *Gahu* and *Soukous*, that are clearly the most dissimilar pair. Furthermore, *Gahu* is the most distinct from all the others.

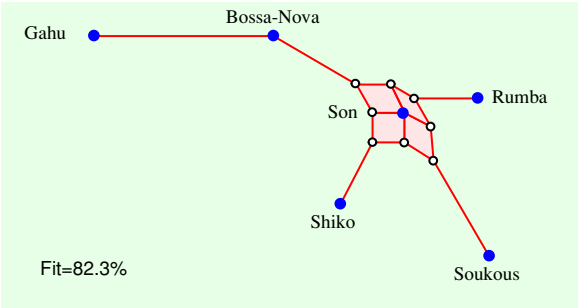

Fig. 13. The SplitsTree with the Euclidean interval vector distance.

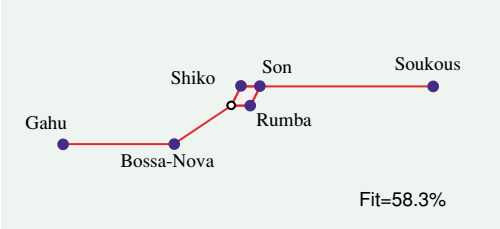

Fig. 14. The SplitsTree with the interval-difference vector distance.

5.1.3 The interval-difference vector distance

For the interval-difference vector distance the splits graph shown in Fig. 14 is almost a chain with a rather poor fit of 58.3%. There is a tight cluster of rhythms comprised of the *Son*, *Rumba*, and *Shiko* patterns. These are the three rhythms that contain identical second bars (the last two beats) given by [. . x . x . . .]. Here also *Gahu* and *Soukous* are the undisputed furthest pair of rhythms.

5.1.4 The swap distance

For the swap distance the splits graph shown in Fig. 15 is a tree with a perfect fit of 100%, and it contains no ancestral nodes. The *Son* is the center of the tree and minimizes both the maximum as well as the sum of the swap distances to all the other rhythms. Therefore the *Son* is both the *Set Swap Center* and the *Set Swap Median* of the set. *Gahu* and *Soukous* are again the unique most dissimilar pair. This splits graph is the simplest of all the graphs.

5.1.5 The chronotonic distance

For the chronotonic distance the splits graph shown in Fig. 16 is a tree-like graph with a perfect fit of 100%. This distance measure is the only one that exhibits a strong clustering in this family of rhythms. In particular, the long parallelogram clearly separates *Bossa-Nova* and *Gahu* from the rest, and invites one to search for a recognizable structure that explains how these two rhythms differ from the others. An examination of their chronotonic

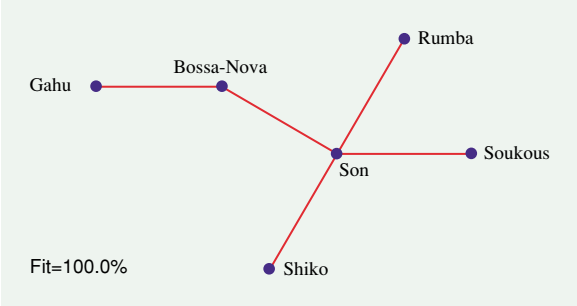

Fig. 15. The SplitsTree with the swap distance.

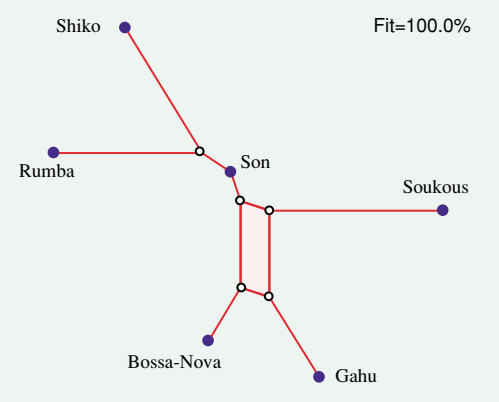

Fig. 16. The SplitsTree with the chronotonic distance.

functions of time (see Fig. 10) yields a clearly distinguishable discriminating feature. The *Bossa-Nova* and *Gahu* functions have only one local maximum, whereas each of the four other rhythms has two local maxima.

5.2. Ternary rhythms

5.2.1 The Hamming distance

For the Hamming distance the splits graph illustrated in Fig. 17 is a complex network with a weak fit of 70.3%. The main observation here is that there is one tightly knit group of five rhythms (located in the upper right) consisting of *Soli*, *Tambú*, *Bembé*, *Yoruba* and *Asaadua*, that are separated from the remaining five which lie scattered about. A discussion of the properties of these five rhythms that sets them apart from the others is deferred to a following section.

5.2.2 The Euclidean interval vector distance

For the Euclidean interval vector distance the splits graph illustrated in Fig. 18 is tree-

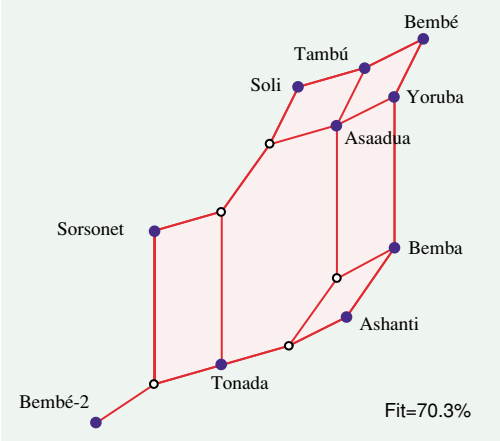

Fig. 17. The SplitsTree with the Hamming distance.

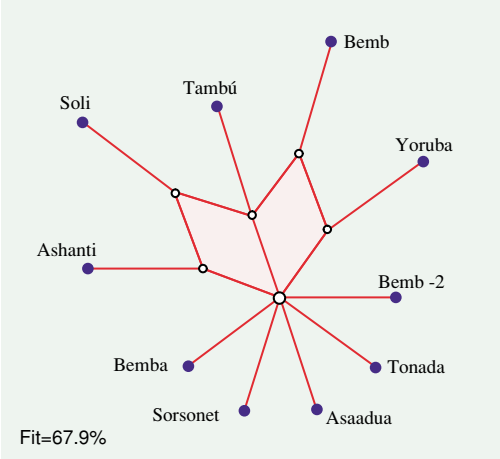

Fig. 18. The SplitsTree with the Euclidean interval-vector distance.

like with a weak fit of 67.9%. Indeed, it is almost a star without any strong clustering evidence. The two parallelograms weakly suggest that *Soli*, *Tambú*, and *Bembé* form one cluster. Analysis of the chronotonic functions in Fig. 11 reveals that these three rhythms have a common feature absent in the other seven, namely, all three end with a short onset interval.

5.2.3 The interval-difference vector distance

For the interval-difference vector distance the splits graph illustrated in Fig. 19 has the lowest fit of all, a meager 28.3%. Yet its two parallelograms have the same structure as

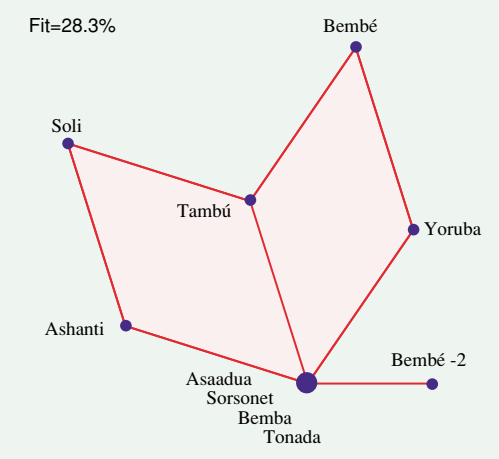

Fig. 19. The SplitsTree with the interval-difference vector distance.

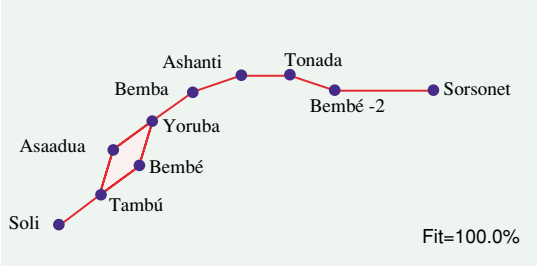

Fig. 20. The SplitsTree with the swap distance.

those in Fig. 18 with the leaves pruned off. It also suggests the same weak clustering of the *Soli*, *Tambú*, and *Bembé* rhythms.

5.2.4 The swap distance

For the swap distance the splits graph illustrated in Fig. 20 is almost a chain with a perfect fit of 100%. There are no ancestral nodes, and the two most dissimilar rhythms, *Soli* and *Sorsonet*, are strikingly visible. The parallelogram (a rhombus) indicates two weak clusters, one comprised of *Soli*, *Tambú*, and *Bembé*, the same as in Figs. 18 and 19, and the other made up of *Soli*, *Tambú*, and *Asaadua*. Examination of the chronotonic functions reveals that these three rhythms are the only ones that have both short inter-onset intervals present in the second halfs of their time spans. Note that the subgraph consisting of the rhombus rhythms (*Tambú*, *Bembé*, *Yoruba* and *Asaadua*) together with *Soli* are the same cluster aparent in the upper right-hand portion of Fig. 17.

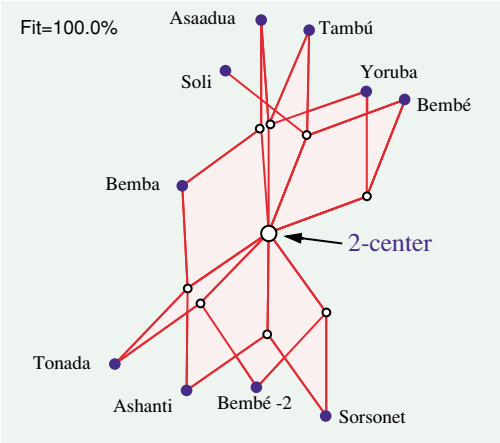

Fig. 21. The SplitsTree with the chronotonic distance.

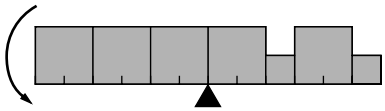

Fig. 22. In the *Soli* rhythm "gravity" rotates the object in a counter-clockwise direction.

5.2.5 The chronotonic distance

For the chronotonic distance the splits graph shown in Fig. 21 exhibits a rich structure with a perfect fit of 100%. There are two highly clustered groups. The first group (already seen in Figs. 20 and 17) consists of *Soli*, *Asaadua*, *Tambú*, *Yoruba*, and *Bembé*. The second group is comprised of *Tonada*, *Ashanti*, *Bembé-2*, and *Sorsonet*. Furthermore, Bemba lies half-way between the two groups.

An analysis of the chronotonic functions in Fig. 11, with the above clustering in mind, yields the following property that explains the clustering. Imagine that each function is a solid object with homogeneous mass equal to the area under the curve, balanced on a pivot located at the midpoint of the time span. For example, Fig. 22 illustrates the *Soli* object balanced on the midpoint pivot. Note that this object has more mass on the left side of the pivot and, by "gravity", would thus rotate in a counter-clockwise direction. So would every other member of group one. On the other hand, all the rhythms of group two have more mass on the right side of the pivot and would rotate in a clockwise direction. Furthermore, the *Bemba* rhythm (which falls half way between the two groups) is perfectly balanced on the pivot.

The splits graph also has nine potential ancestral nodes. One of these is quite distinguished from the rest, not only because of its high degree, but because it is located

in the center of the entire collection, and is at a distance of two from all ten rhythms. For this reason it is labelled the *2-center* of the set. A reconstruction of this ancestral rhythm yields the pattern [x . x . x . x . x . x .], i.e., a steady isochronous "heart-beat" consisting of six beats per measure.

6. Concluding Remarks

Several desirable criteria spring to mind for comparing phylogenetic trees of families of rhythms. Such criteria include (1) simplicity, (2) goodness of fit, (3) indication of clustering, (4) generation of interesting "ancestral" rhythms, and (5) support of musicological analyses.

Simplicity: The simplest splits graphs are obtained with the swap distance. For the binary rhythms the splits graph is a tree, and for the ternary rhythms there is only one cycle. Furthermore, neither graph has "ancestral" nodes.

Goodness of fit: The only distance measures that yield a 100% fit for both binary and ternary rhythms are the swap and chronotonic distances. The Hamming distance also produced a 100% fit for the binary rhythms.

Clustering: The most impressive clustering is obtained with the chronotonic distance. For the binary rhythms, Gahu and Bossa-Nova are clearly differentiated from the rest. For the ternary rhythms the splits graph yields two well separated clusters.

Ancestral rhythm generation: The most noteworthy "ancestral" rhythms were produced with the chronotonic distance for ternary rhythms and with the Hamming distance for the binary rhythms.

Musicological analysis: According to David Locke (LOCKE, 1998) *Gahu* is one of the most unusual binary bell patterns, and he devotes a lengthy musicological analysis to make this point. His conclusions are well supported by the results obtained here with the Euclidean, swap, intervaldifference, and chronotonic distance functions.

In conclusion, a comparison of the significance of the role played by the five distance measures according to each of the five criteria outlined in the preceeding, suggests that, in this context, the best overall rhythmic dissimilarity measure is the chronotonic distance, followed by the swap distance in close second place.

With respect to computational complexity, all five measures may be computed efficiently in $O(n)$ time, where n is the length of the binary representation in which the rhythm is given.

I thank David Bryant for discussions on the *SplitsTree* program, Pat Morin and Tom Shermer for illuminating discussions on the swap distance, and Kjell Gustafson for references on the history of his chronotonic (TEDAS) representation of rhythm.

REFERENCES

ALOUPIS, G., FEVENS, T., LANGERMAN, S., MATSUI, T., MESA, A., NUÑEZ, Y., RAPPAPORT, D. and TOUSSAINT, G. (2003) Computing a geometric measure of the similarity between two melodies, *Proc. 15th Canadian Conf. Computational Geometry*, pp. 81–84, Dalhousie University, Halifax, Nova Scotia, Canada, August 11–13.

AROM, S. (1991) *African Polyphony and Polyrhythm*, Cambridge University Press, Cambridge, England.

BERENZWEIG, A., ELLIS, D. P. W. and LAWRENCE, S. (2003) Anchor space for classiffication and similarity measurement of music, *Proc. ICME-03*, Baltimore, U.S.A., July.

BIEDL, T., CHAN, T., DEMAINE, E. D., FLEISCHER, R., GOLIN, M. and MUNRO, J. I. (2001) Fun-Sort—Or the chaos of unordered binary search, *Discrete Applied Mathematics*, **144**, 231–236.

BOOKSTEIN, A., KULYUKIN, V. A. and RAITA, T. (2002) Generalized Hamming distance, *Information Retrieval*, **5**(4), 353–375.

CHA, S.-H. and SRIHARI, S. N. (2002) On measuring the distance between histograms, *Pattern Recognition*, **35**, 1355–1370.

COYLE, E. J. and SHMULEVICH, I. (1998) A system for machine recognition of music patterns, *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Seattle, Washington, D.C.

DE BRUIJN, N. G. (1974) Sorting by means of swapping, *Discrete Mathematics*, **9**, 333–339.

DRESS, A., HUSON, D. and MOULTON, V. (1996) Analysing and visualizing sequence and distance data using SPLITSTREE, *Discrete Applied Mathematics*, **71**, 95–109.

EKWUEME, L. E. N. (1974) Concepts in African musical theory, *Journal of Black Studies*, **5**(1), 35–64.

FOOTE, J., COOPER, M. and NAM, U. (2002) Audio retrieval by rhythmic similarity, *Proc. Third International Conference on Music Information Retrieval*, Centre Pompidou, Paris, France, October 13–17.

GASIENIEC, L., JANSSON, J. and LINGAS, A. (1999) Efficient Approximation algorithms for the Hamming center problem, *Proc. 10th Annual ACM-SIAM Symposium on Discrete Algorithms*, S905–S906.

GOLDROSEN, J. (1975) *The Buddy Holly Story*, The Bowling Green University Popular Press, New York.

GUSTAFSON, K. (1988) The graphical representation of rhythm, *(PROPH) Progress Reports from Oxford Phonetics*, University of Oxford, **3**, 6–26.

HAMMING, R. W. (1986) *Coding and Information Theory*, Prentice-Hall, Englewood Cliffs. HOFMANN-ENGL, L. (2002) Rhythmic similarity: A theoretical and empirical approach, *Proc. Seventh International Conference on Music Perception and Cognition*, pp. 564–567, Sidney, Australia.

HOFMANN-ENGL, L. (2002) Rhythmic similarity: A theoretical and empirical approach, *Proc. Seventh International Conference on Music Pereption and Coguition*, Sidney, Australia, 564–567.

HUSON, D. H. (1998) SplitsTree: Analyzing and visualizing evolutionary data, *Bioinformatics*, **14**, 68–73.

JIANG, X., ABEGGLEN, K., BUNKE, H. and CSIRIK, J. (2003) Dynamic computation of generalized median strings, *Pattern Analysis and Applications*, **6**, 185–193.

KRUSKAL, J. B. (1983) An overview of sequence comparison: time warps, string edits, and macromolecules, *SIAM Review*, **25**(2), 201–237.

LOCKE, D. (1998) *Drum Gahu: An Introduction to African Rhythm*, White Cliffs Media, Gilsum, New Hampshire.

MATHAI, A. and RATHIE, P. (1975) *Basic Concepts in Information Theory and Statistics*, Wiley Eastern Ltd., New Delhi.

MEUDIC, B. (2003) Musical similarity in a polyphonic context: a model outside time, *Proc. XIV Colloquium on Musical Informatics*, Firenze, Italy, May 8–10.

MOLES, A. (1966) *Information Theory and Esthetic Perception*, The University of Illinois Press, Urbana and London.

MONGEAU, M. and SANKOFF, D. (1990) Comparison of musical sequences, *Computers and the Humanities*, **24**, 161–175.

MONT-REYNAUD, B. and GOLDSTEIN, M. (1985) On finding rhythmic patterns in musical lines, *Proc. International Computer Music Conference*, 391–397.

ÓMAIDÍN, D. (1998) A geometrical algorithm for melodic difference, *Computing in Musicology*, **11**, 65–72.

ORPEN, K. S. and HURON, D. (1992) Measurement of similarity in music: A quantitative approach for non-parametric representations, *Computers in Music Research*, **4**, 1–44.

RAHN, J. (1987) Asymmetrical ostinatos in sub-saharan music: time, pitch, and cycles reconsidered, *In Theory Only*, **9**(7), 23–37.

RAHN, J. (1996) Turning the analysis around: African-derived rhythms and Europe-derived music theory, *Black Music Research Journal*, **16**(1), 71–89.

SHMULEVICH, I., YLI-HARJA, O., COYLE, E., POVEL, D.-J. and LEMSTRÖM, K. (2001) Perceptual issues in music pattern recognition, *Computers and the Humanities*, **35**, 23–35.

TOUSSAINT, G. T. (1974a) On the divergence between two distributions and the probability of misclassiffication of several decision rules, *Proc. Second International Joint Conference on Pattern Recognition*, pp. 27–35,

Copenhagen, Denmark, August 13–15.

TOUSSAINT, G. T. (1974b) On some measures of information and their application to pattern recognition, *Proc. Conference on Measures of Information and their Application*, pp. 21–28, Bombay, India, August 16–18.

TOUSSAINT, G. T. (1975) Sharper lower bounds for discrimination information in terms of variation, *IEEE Transactions on Information Theory*, January, 99–100.

TOUSSAINT, G. T. (2002) A mathematical analysis of African, Brazilian, and Cuban clave rhythms, *Proc. BRIDGES: Mathematical Connections in Art, Music and Science*, pp. 47–56, Towson University, Towson, MD, U.S.A., July 27–29.

TOUSSAINT, G. T. (2003a) Classification and phylogenetic analysis of African ternary rhythm timelines, *Proc. BRIDGES: Mathematical Connections in Art, Music and Science*, pp. 47–56, Granada, Spain, July 23–27.

TOUSSAINT, G. T. (2003b) Algorithmic, geometric, and combinatorial problems in computational music theory, *Proc. X Encuentros de Geometria Computacional, University of Sevilla*, pp. 101–107, Sevilla, Spain, June 16–17.

TYPKE, R., GIANNOPOULOS, P., VELTKAMP, R. C., WIERING, F., and VAN OOSTRUM, R. W. (2003) Using transportation distances for measuring melodic similarity, *Proc. Fourth International Conference on Music Information Retrieval (ISMIR)*, 107–114.

WRIGHT, O. (1978) *The Modal System of Arab and Persian Music AD 1250–1300*, Oxford University Press, Oxford, England.