Measuring musical rhythm similarity:
Transformation versus feature-based methods

Godfried T. Toussaint[1], Luke Matthews[2], Malcolm Campbell[3], and Naor Brown[4]

[1] Department of Music, Harvard University
[2] Department of Human Evolutionary Biology, Harvard University
[3] New England Conservatory of Music, Boston
[4] Department of Computer Science, Harvard University

Background in musicology. A central problem in musicology is to develop classification algorithms that match the way human observers perceive musical rhythm similarity and pass on music forms as cultural knowledge. Rhythm similarity measures that have been investigated in the past fall into two categories: transformation methods and feature-based methods. In the former the similarity is measured by the amount of effort required to transform (morph, mutate) one rhythm into another. One of the most well know transformation measures is the *edit* distance, defined as the minimum number of *mutations* required to transform one rhythm to the other. The mutations here are *insertions*, *deletions*, and *substitutions* of symbols in a sequence. In the feature-based methods a set of pre-determined features is calculated for each rhythm, and similarity is measured by the degree to which the two sets of features match. Typical features measure properties of the inter-onset intervals present in a rhythm. For either approach, a measure of rhythm similarity is desired that is not only efficient to compute, but that agrees well with human perceptual judgments of rhythm similarity.

Background in cultural evolutionary biology. Phylogenetic trees were originally conceived for the purposes of describing and visualizing evolutionary relationships that exist between members of a group of biological organisms. However, more recently they have been applied to cultural objects as well, language being an early prime example. Phylogenetic tree construction approaches fall into two main categories: distance-based methods and "character"-based methods. Distance methods assume that a distance matrix is available containing the distance between every pair of objects being studied. In character-based methods the input data are sets of binary and/or multistate characters, and the final observed distribution of characters is modeled as resulting from a set of inferred transition probabilities.

Aims. By combining phylogenetic tools from cultural evolutionary biology with the cognitive and computational study of measures of rhythm similarity from musicology and music theory, this project aims to create a synergistic bridge between these two domains of knowledge.

Main contribution. A feature-based approach to rhythm similarity is compared to a frequently used transformation method, using a family of rhythms of equal length. Distance matrices calculated from the rhythms are compared with 'dissimilarity' matrices obtained from human judgments using Mantel tests. Two different phylogenetic analysis techniques are also compared: a distance based method and a Bayesian approach. The results provide evidence from the music domain that supports the hypothesis that transformation methods are superior to feature-based methods for modeling more general human similarity judgments.

Implications. Our results highlight the difficulty of modeling rhythm similarity by means of purely structural features collected from musicology and music theory, and imply that more attention should be devoted to the study of transformation methods. The methods explored here, standard in evolutionary biology and anthropology, provide novel tools for musicology and music theory that complement traditional historical and ethnographic accounts.

1.   Introduction

Defining and measuring similarity are central concerns in vast areas of scientific and artistic enquiry. Not surprisingly, these problems have been investigated extensively in a wide range of different disciplines, resulting in the exploration of a variety of approaches. Much of the literature deals with similarity in the context of visual perception of patterns in space. This paper on the other hand is concerned with the problem of measuring musical rhythm similarity, that is, auditory patterns in time. For this project rhythms are stripped to their barest representation as pure duration patterns without any form of accents or alternation of timbres (Rivière, 1993). Furthermore the representations of rhythms are symbolic, as in standard Western notation, and therefore they are exact, without any micro-timing variations.

To clarify the main general approaches to measuring similarity in this context, consider the three well-known rhythms, bossa-nova, rumba, and son, illustrated using polygon notation in Figure 1. It is convenient to represent the rhythms as cyclic binary sequences on a clock diagram consisting of sixteen evenly spaced pulses. The sounded pulses (also called onsets) are marked with solid black circles, whereas the silent pulses are white. The rhythms start at pulse 0, and time flows in a clockwise direction. The adjacent onsets are connected with line segments to create convex polygons that aid visualization of the rhythmic features, and the sides of the polygons are labeled with numbers that indicate the durations of adjacent inter-onset-intervals (IOIs).
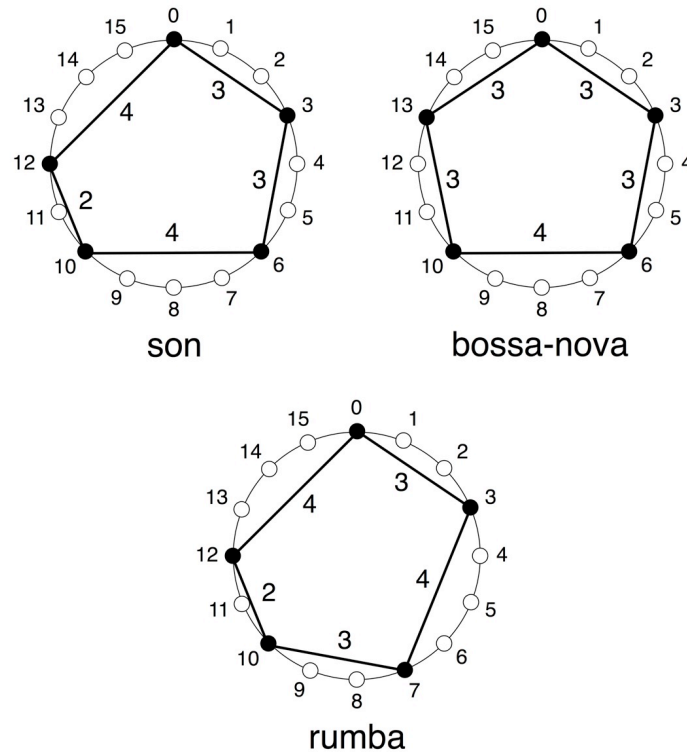

Figure 1. Three rhythms represented in polygon notation on a cyclic clock.

In the first approach the rhythms are analyzed in terms of features that are present or absent in each rhythm. From a list of such features the similarity between two rhythms is measured by a function of the common and distinctive features that they possess. The table in Figure 2 lists, for example, four features that are present or absent in these three rhythms. The first feature indicates whether the rhythm has mirror symmetry about a line passing through two pulse positions. The bossa-nova and son are symmetric with respect to the pulse pairs (0, 8) and (3, 11), respectively. The meanings of the remaining three features are self-evident from their labels. In the absence of other information, these features would affirm that the son is more similar to the bossa-nova than to the rumba, since it shares two features with the former, and only one with the latter. This approach is knows as Tversky's contrast model of similarity (Tversky, 1977).

Features	Son	Bossa-Nova	Rumba
Mirror Symmetry	Yes	Yes	No
IOI Duration Equal to 2	Yes	No	Yes
Adjacent IOIs of Equal Duration	Yes	Yes	No
3-4-3 Duration Pattern	No	Yes	Yes

Figure 2. Table of presence and absence of four features in the three rhythms of Figure 1.

A second approach to measuring similarity makes a set of d measurements (variables, features) on the objects of study, thereby mapping the objects to points in a d-dimensional feature space. Two objects are deemed similar in this scheme if the distance (according to some metric or more general measure) between their corresponding points in feature space is relatively small. An example of a 2-dimensional feature space for the three rhythms of Figure 1 is pictured in Figure 3, where the horizontal axis measures the longest sequence of adjacent inter-onset intervals (in terms of the number of intervals) present in the rhythm, and the vertical axis indicates the duration of the rhythm's shortest inter-onset interval. If the distances between these three points are measured with the Euclidean distance, for example, then the son would be considered to be more similar to the rumba than the bossa-nova, in contradiction with Tversky's contrast model. This method belongs to the spatial approach of Shepard, who measured distances in a psychological space (Shepard, 1957). This approach has had a profound influence on the fields of automated pattern recognition and information retrieval, where all manner of analogous mathematical feature spaces and distance metrics are routinely employed (Duda, Hart & Stork, 2000; Nilsson, 1990).
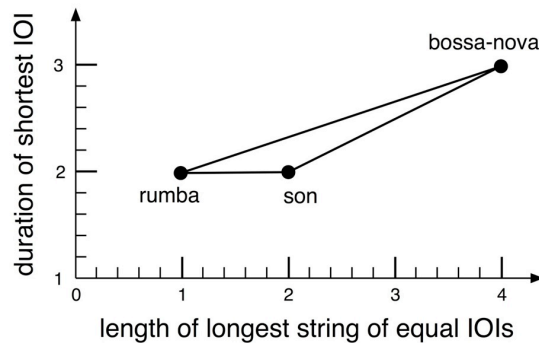

Figure 3. A 2-dimensional feature space for the rhythms in Figure 1.

The two aforementioned approaches have been criticized on the grounds that their representations of the objects to be compared are arbitrary and over-simplistic (Hahn & Chater, 1997; Hahn & Chater, 1998a; Hahn & Chater, 1998b). The authors of these papers argue that complex patterns require structural representations that specify not just *what* features are to be used, but *how* they are interrelated.

The third approach to measuring the similarity between two objects bypasses the creation of lists of features, and instead operates directly on the structure of the objects themselves. First a group of fundamental operations, also called edits or mutations, is established, by which small changes may be made to the structure of the objects under consideration. The distance between two objects is then defined to be the minimum number of such operations that must be made on one object in order to *transform* it to the other. In the case of the three rhythms in Figure 1, for instance, one of the simplest changes that may be made is to move an onset from one pulse position to either of its two adjacent neighboring positions. Such a change is called a *swap*, and the minimum number of swaps required to transform one rhythm to another is termed the *swap distance* between them (Toussaint, 2002; Toussaint, 2004; Lowrance & Wagner, 1975). From Figure 1 it may be observed that the son may be converted to the bossa-nova by swapping the onset at pulse 12 with the silent pulse 13, and to the rumba by swapping the onset at pulse 6 with the silent pulse 7. Thus in both cases the swap distance is 1, implying that by this reckoning the son is *equally* similar to the bossa-nova and the rumba, in contradiction to *both* the contrast model of Tversky and the feature-space model of Shepard.

Transformation methods have been used extensively to measure the similarity between symbol sequences in a variety of areas such as text processing (Lowrance & Wagner, 1975), computational linguistics (Wieling, 2007), bioinformatics (Gusfield, 1997; Sankoff & Kruskal, 1999), musicology, (Crawford, Iliipoulos, & Raman, 1997-98), and music information retrieval (Typke, Veltkamp & Wiering, 2004). Furthermore, recent experiments in the visual domain have shown that distances based on transformations have a strong impact on similarity judgements, and indeed challenge the feature-based methods (Hahn, Chater, & Richardson, 2003; Hodgetts, Hahn, & Chater, 2009). Although in the music domain there is evidence that the edit distance (also known as the Levenshtein distance), a transformation method popular in music information retrieval applications, is a good predictor of human perceptual judgments (Toussaint, Malcolm & Brown, 2010a), no previous comparisons with feature-based methods have been reported. Here a feature-based approach to measuring musical rhythm similarity is compared to the edit-distance (which is more powerful than the swap distance). These two methods are compared to each other with respect to human judgments of perceived similarity by means of two different phylogenetic analysis techniques: a distance based method and a Bayesian approach. While these techniques are standard in the field of evolutionary biology, the procedures are novel in the area of music perception, and it is hoped their introduction here will showcase their applicability. Furthermore, the results obtained add evidence from the aural domain to that already collected from the visual domain, to support the hypothesis that transformation methods may be superior to feature-based methods for modeling human perception of similarity.

2.   The Feature-Based Approach

A feature-based approach to similarity poses several problems outright. First among these is the fact that there are an infinite number of possible features. For example, one feature might be: the rhythm contains less than six onsets. A second feature might

be: the rhythm contains less than seven onsets. Furthermore, the number seven may be replaced with any integer greater than seven to generate an infinite number of additional features. How many features should be chosen? Once a number has been decided upon, which features should be selected from this distinguished set? Even for visual patterns selecting the variables that produce grouping by similarity can be difficult (Olson & Attneave, 1970). One of the goals of the project described here is not to determine what features produce similarity grouping, a formidable task, but rather the much more modest one of determining if the structural features that are commonly used in musicology and music theory to categorize and characterize musical rhythms do produce significant similarity groupings. In this section fourteen features are described that measure a variety of structures of cyclic rhythms, and that were chosen on the basis of concepts considered important in theories of musical rhythm or properties of rhythm used in the classification of music styles.

1st.   Number of pulses is an even number

The number of pulses in the cycle of a musical rhythm is a feature that distinguishes music from different parts of the world (Toussaint, 2005). For instance, much folk dance music of Rumania has cycles of 5, 7, and 9 pulses (Proca-Ciortera, 1969), whereas much drum music from Sub-Saharan Africa uses cycles with 12, 16 and 24 pulses (Arom, 2004). Therefore rhythms may be usefully categorized as having cycles containing an odd or even number of pulses.

2nd.   Number of onsets is an even number

The parity of the number of onsets in a rhythm does not appear to be as important musically as the parity of the number of pulses, and for large numbers of onsets in a cycle the parity may not have much relevance in categorization or perception. Often, adding decorative onsets does not markedly change the perception of the rhythm, although the effect may depend on where in the cycle the onset is added. However, when the number of onsets is small, rhythms with an odd number of onsets sound sufficiently different from those with an even number, to warrant the inclusion of this feature in the set investigated. For example, for the special rhythms with 12 and 16 pulses, there is a marked preference for 5 and 7 onsets, respectively (Toussaint, 2003).

3rd.   Number of onsets is greater than half the number of pulses

The number of notes in a phrase of music, referred to as the *phrase density* (also note density), has been used as a rhythmic feature in the systematic statistical analysis of music (Cerulo, 1998), and has been found to be a highly reliable predictor of melodic coherence (Brown, Towsey, Wright, & Deiderich, 2001). The number of tones per second, called the *rhythmic activity*, has also been used as a rhythmic feature in comparisons with perceived similarity of melodies (Eerola, Järvinen, Louhivuori, & Toiviainen, 2001). If the timespans or phrase-lengths of the rhythms are normalized for time, then the rhythmic activity and phrase density are equivalent features. The phrase density of a rhythm is closely related to its *rhythmicity*: the number of transitions between a note and a pause (Van Den Broek & Todd, 2009-10). This number is a measure of the rhythm's *isochrony*. A rhythm may contain many or few onsets relative to the number of pulses in its periodic cycle. The former will be called *dense*, and the latter *sparse*. A rhythmically important boundary value to distinguish between dense and sparse rhythms is half the number of pulses (almost half if the number of pulses is odd). The most popular rhythm timelines composed of twelve and

sixteen pulses have 7 and 5 onsets, respectively, close to one half their pulse numbers (Toussaint, 2002; Toussaint, 2003). Figure 4 shows a sparse rhythm (the tresillo) popular in Cuba (left) and a dense Tuareg rhythm from Lybia (right) (Toussaint, 2005). In general, sparse rhythms sound quite different from dense rhythms.
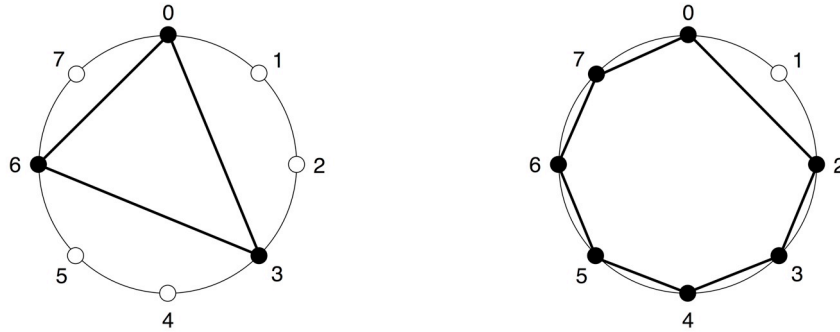
Figure 4. A sparse rhythm (left) and a dense rhythm (right).

4th.   First onset is not on first pulse (anacrusis)

In European common practice music the most accented onset (the main downbeat) occurs at pulse 0, the start of the rhythmic cycle (Lezcano, 1991; Temperley, 2010). On the other hand, in much of Sub-Saharan African music this onset is often missing from the rhythm (Chemillier & Truchet, 2003). This property is called the *silent down beat* or anacrusis, a feature that is present in some flamenco metric patterns of Southern Spain (Guastavino, et al., 2009), and absent in traditional Korean music (Howard, 1992). Figure 5 shows two Afro-Cuban rhythms with anacrusis: the reverse clave son (left) and the guaguancó (right). The property of anacrusis affects the perception of a rhythm or melody. In the words of Justin London (London, 2009), "When a melody begins with an anacrusis, rhythmic grouping structure and meter are out of phase." Although none of the rhythms in the data used for the present experiments have anacrusis, the feature was included for future experiments that will incorporate rhythms that have this property.
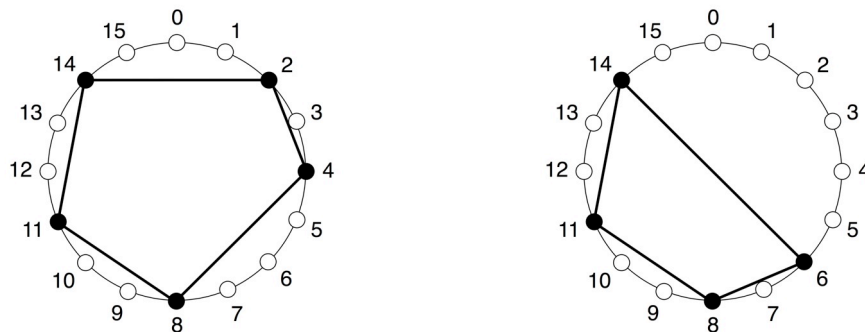
Figure 5. Two Afro-Cuban rhythms exhibiting anacrusis: the 2-3 clave son (left) and the guaguancó (right).

5th.   Evenness

A property that has received a great deal of attention in music theory in both the pitch and time domains is the concept of *maximal evenness* (Domínguez, Clampitt, & Noll, 2009; Amiot, 2007; Clough & Douthett, 1991). Intuitively speaking, a rhythm is maximally even if the onsets are distributed within the cycle as evenly as possible. A regular isochronous rhythm is *perfectly even*. If the number of onsets in a maximally even rhythm divides without remainder into the number of pulses, then the rhythm is always perfectly even. There are several different characterizations of maximally even sets. One of these states that a rhythm is maximally even if, and only if, the sum of all the pair-wise straight-line distances between the corresponding points on the circle is maximized (Toussaint, 2005a). The ancient Euclidean algorithm for determining the greatest common divisor of two numbers may also be used to generate maximally even rhythms (Toussaint, 2005b). Figure 6 (left) shows a maximally even rhythm used in samba music with seven onsets in a sixteen-pulse cycle. The two methods described above are relatively complicated. A much simpler algorithm for generating maximally even rhythms with k onsets in a cycle of n pulses constructs a regular polygon of k sides on the circle containing n lattice points, and then snaps (moves) all the vertices of the regular polygon that do not lie on a lattice point to their nearest clockwise lattice point. For example, consider the case when $n = 16$ and $k = 7$. The locations of the vertices of a regular 7-sided polygon (heptagon) may be obtained by dividing 16 by 7 yielding an inter-point distance of 2.286 to obtain the regular heptagon with vertices at locations 0, 2.268, 4.536, 6.804, 9.072, 11.34, and 13.608. Rounding up these numbers that are not integers to their nearest clockwise integer yields 0, 3, 5, 7, 10, 12, and 14, with resulting inter-onset interval pattern 3-2-2-3-2-2-2, which is a rotation of the rhythm in Figure 6 (left) with interval pattern 2-3-2-2-3-2-2. All rotations of a maximally even pattern are of course also maximally even.

If each polygon vertex is permitted to be snapped in either direction (clockwise or counterclockwise) then a larger set of rhythms is generated called *almost maximally even* (Toussaint, 2011). The seven-onset rhythm in the center of Figure 6, a binary version of the ternary standard pattern (Agawu, 2006) is not maximally even, but is almost maximally even. Finally, the rhythm on the right is neither maximally even nor almost maximally even. Accordingly, the evenness feature takes on the value 2 if the rhythm is maximally even, 1 if it is almost maximally even, and 0 otherwise.
Symmetry in music, especially mirror (reflective) symmetry, has played a key role in the aesthetics, analysis, and compositional practice of music (Kempf, 1996; Dean, Byron, & Bailes, 2009-10; Porter, 1970; Wilson, 1986; Liebermann & Liebermann, 1990; Herrero, 2010). Indeed, Simha Arom has suggested that symmetry might be one of the *universals* of music (Arom, 2001). Wilfrid Hodges documents the variety of symmetries in music that have been used extensively by composers during the past seven centuries (Hodges, 2003). The features numbered 6 through 9 measure four types of mirror symmetry.

6th.   Vertical mirror symmetry
7th.   Horizontal mirror symmetry
8th.   Positively skewed mirror symmetry
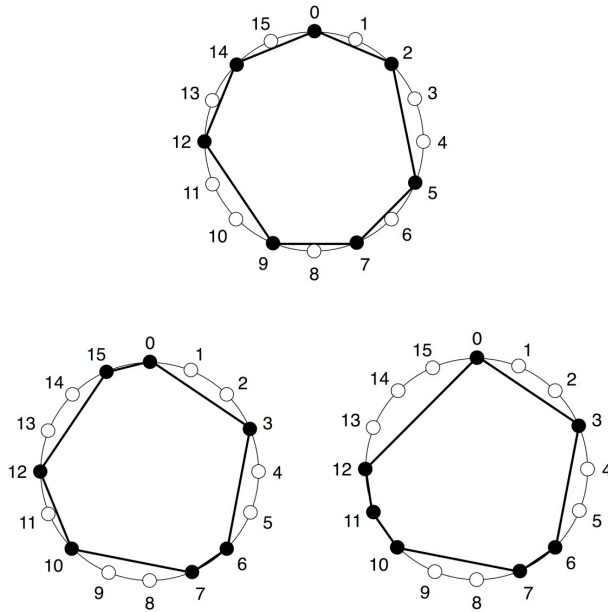**9th.   Negatively skewed mirror symmetry**

Figure 6. Maximally even rhythm (left) almost maximally even rhythm (center), and non-maximally even rhythm (right).
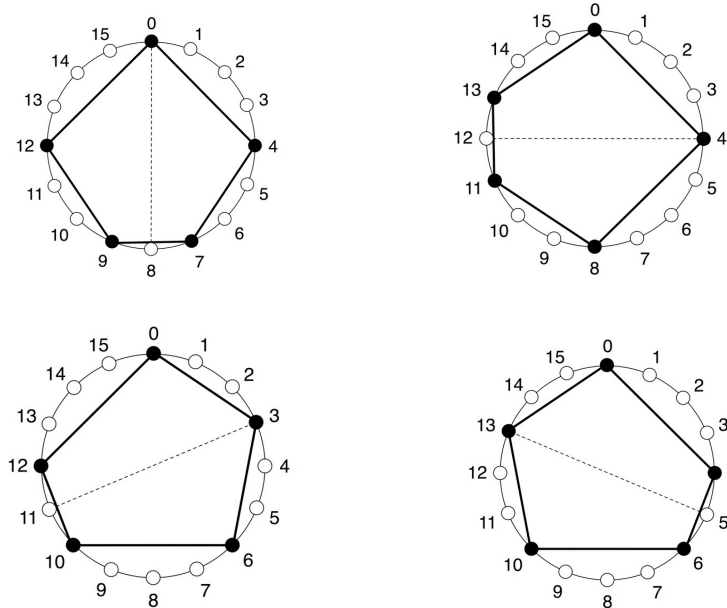

Figure 7. The four types of symmetry: vertical (upper left), horizontal (upper right), positively skewed (lower left), and negatively skewed (lower right).

These four types of symmetry are illustrated in Figure 7. Although these features are easy to interpret visually, the more relevant issue in the present context is what they mean musically to a listener. Musically, a 16-pulse timeline in 4/4 time has four fundamental beats at pulses 0, 4, 8, and 12, and hence the vertical and horizontal symmetries reflect this metrical division of time. Equally important in much contemporary and traditional world music is the breaking of symmetry present in rhythms (Feldman, 1981). For example, the asymmetric *aksak* rhythms form a distinctive feature of Balkan folkloric rhythms (Fracile, 2003), as do the asymmetric *timeline* bell patterns in West African music (Magill & Pressing, 1997). The positively and negatively skewed axes of symmetry measure the presence of this type of asymmetry. Whether it is possible for the visual symmetries of polygons to be perceived by listeners when the corresponding rhythms are realized acoustically remains an interesting topic for future research.

10th.   Antipodality

A rhythm has the *antipodal* property if when represented on a circle it contains two onsets that lie diametrically opposite each other. For instance, the door-knock rhythm in Figure 8 (left) has this property because there are onsets at pulses 4 and 12, which determine a diameter of the circle. On the other hand, the clave son (right) does not have this property. A rhythm with this property places emphasis on the division of the time-span of the rhythm cycle into two half cycles of equal duration, and thus creates some regularity. A rhythm that does not possess this property is said to have the *rhythmic oddity* property, a terminology introduced by Simha Arom (Arom, 2004; Chemillier & Truchet, 2003). The antipodality feature takes on three values: 0 if there is no antipodal pair, 1 of there is exactly one antipodal pair, and 2 if the rhythm admits more than one antipodal pair. The number of antipodal pairs, or the amount of rhythmic oddity that a rhythm possesses has been used to categorize Sub-Saharan African timelines (Toussaint, 2005). Note that this feature is not binary. It was felt that more discrimination information would be captured by making the feature take on three values.
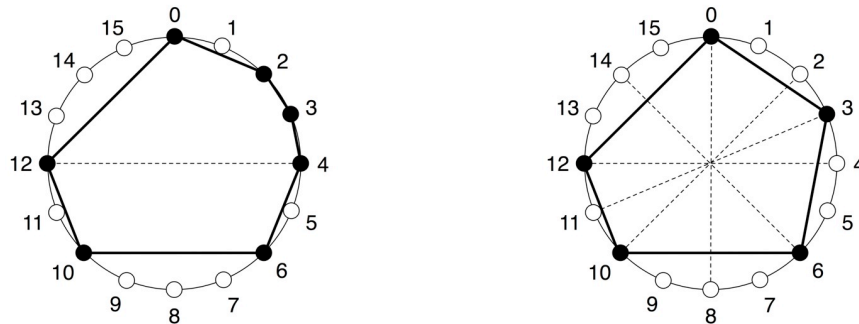

Figure 8. A rhythm with the antipodal property (left) and one without (right).

11th.   Off-Beatness

Consider the 16-pulse cycle pictured in Figure 9 (left). The number sixteen may be evenly divided (without remainder) by only three numbers that are smaller than

sixteen and greater than one, namely 8, 4, and 2. These are the only possible regular rhythms that may be embedded in the 16-pulse cycle, and they occupy all the pulses other than 1, 3, 5, 7, 9, 11, 13, and 15. These are all the numbers that are relatively prime to 16, and therefore indicate the positions in the cycle where the presence of an onset would be considered to be off the beat in the context of music that uses regular rhythms. Onsets at these positions may thus be considered to possess a kind of syncopation that may be called *off-beatness* (Toussaint, 2005). The off-beatness measure of a rhythm is the sum of the number of onsets it contains that lie on these off-beat positions. For example, the clave son rhythm in the center has an off-beatness value of one due to its onset at pulse 3, whereas the clave rumba on the right has an off-beatness value of two by virtue of its onsets at pulses 3 and 7. The off-beatness feature takes on three values: 0 if there are no off-beats, 1 if the number of off-beat onsets is one or two, and 2 if there are more than two off-beats. It has been shown empirically that the off-beatness measure is closely related to several mathematical measures of syncopation as well as human judgments of meter complexity (Thul & Toussaint, 2008a). Furthermore, syncopation has been demonstrated to be a highly reliable predictor of melodic coherence (Brown, Towsey, Wright, & Deiderich, 2001), thus motivating the use of the off-beatness feature. Note that this feature is also not binary. It was also felt that more discrimination information would be captured by making this feature take on three values.
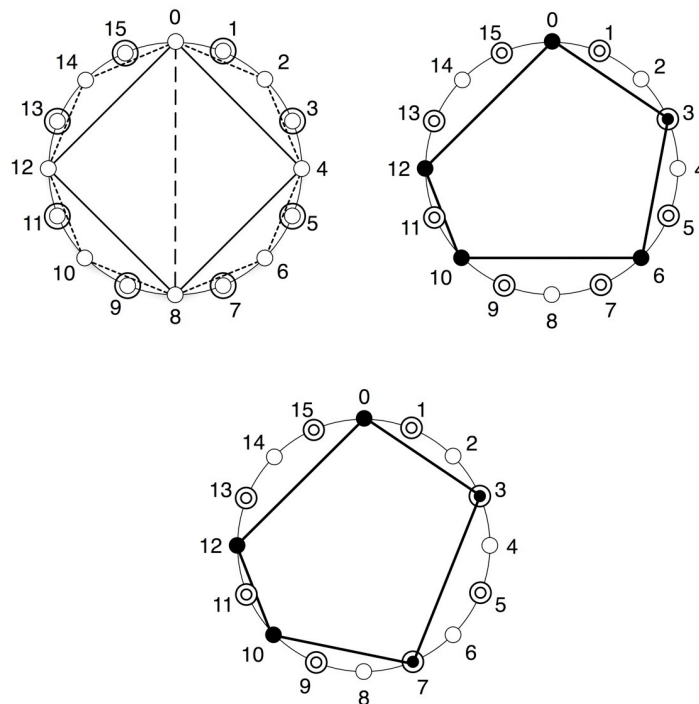

Figure 9. The off-beat positions (left), a rhythm with off-beatness = 1 (center) and one with off-beatness = 2 (right).

12th.   Hemiola property

Hemiola is the name given to a property that characterizes the metric patterns of some Sub-Saharan African, Afro-Cuban, and flamenco music of southern Spain. The traditional African hemiola (Brandel, 1959) has the metric pattern [3-3-2-2-2] shown in Figure 10. In this form it divides the 12-pulse cycle into two half-cycles of six pulses each, and then divides the first half-cycle into two equal 3-pulse intervals, and the second half-cycle into three equal 2-pulse intervals. This property may be generalized to rhythms with time-spans of cardinalities other than twelve, as long as the two half-cycles of the rhythm admit divisions into regular rhythms having intervals of different lengths, such as the two examples illustrated in Figure 11.
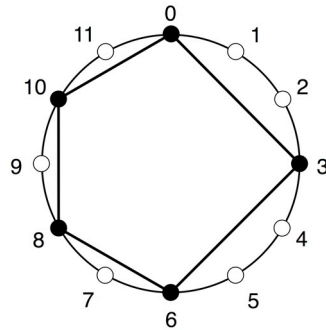
Figure 10. The traditional African style hemiola.
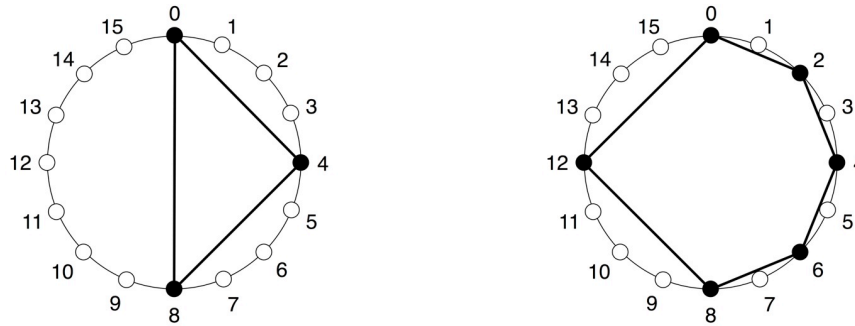
Figure 11. Generalized Hemiola property with 3 onsets (left) and 6 onsets (right).

13th.   Toggle property

A popular hand-drumming technique strikes the drum on all the pulses of a rhythm very gently, alternating with the right and left hands. The rhythm emerges from accenting in some way those pulses where onsets should occur. Consider the two rhythms in Figure 12, where each pulse is labeled with the letters R and L to denote striking the drum with the right and left hands, respectively. The clave son (left) has a hand-pattern given by RLRRR, whereas the samba variant (right) has the hand-

pattern RLRLR. A rhythm has the *toggle* property if the hand-pattern cycle contains only one transition from R to L and one from L to R (Toussaint, 2010). In other words, toggle rhythms are those cyclic rhythms that when played by alternating the hands on each pulse, have their onsets divided into two consecutive sets, such that the first set is played consecutively with one hand, and subsequently the second set is played consecutively with the other hand. Playing this way feels as if one hand responds to a question posed by the other hand, analogous to the customary call-and-response method of singing existent in much of Sub-Saharan Africa. Thus the clave son has the toggle property, whereas the samba variant does not. An alternate geometrical interpretation is possible when the rhythms are represented on a circle, as in Figure 12. A rhythm has the toggle property if, and only if, there exists a straight line that separates the R onsets from the L onsets. The clave son rhythm on the left admits a separating line determined by pulses 1 and 5. On the other hand, for the samba variant on the right no separation line exists. Most of the rhythmic timelines used in Sub-Saharan music are distinguished by having this property (Rahn, 1987; Rahn, 1996). Note that the toggle property is weakly related to the off-beatness property. In a 16-pulse timespan the odd-numbered pulses become off-beat positions. Therefore a rhythm with the toggle property is guaranteed to have some off-beat onsets. However, the converse is not true, as the rhythm in Figure 12 (right) shows: it has two off-beat onsets at pulses 3 and 9, and yet it does not have the toggle property.
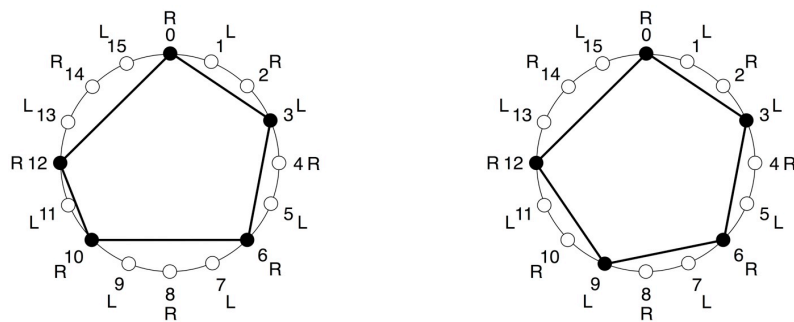

Figure 12. Rhythm with the toggle property (left) and without (right).

14th.   Balanced rhythms

A rhythm is *balanced* if every line that passes through the center of the circle partitions the onsets of the rhythm into two sets with cardinalities that differ by at most one (Jiang, 2008). Consider the two rhythms in Figure 13. The clave son rhythm on the left is balanced. The vertical line through pulses 0 and 8 leaves two onsets on each side, and the line through pulses 1 and 9 separates two onsets from three. However, no line leaves one side with four or five onsets. On the other hand, the famous *door-knock* rhythm on the right contains a line through pulses 0 and 8 that has four onsets on one side and two on the other, and therefore it is not balanced. The property of being balanced arises in the context of the theory of maximally even sets (Clough & Douthett, 1991). A property of musical chords studied by Douthett and Entringer (Clough & Douthett, 1991, note 15 on p. 40) maximizes the sum of the

pair-wise arc lengths between the chord's pitches represented as points on a circle. It turns out that this sum is maximized if, and only if, the set of points is balanced (Jiang, 2008). The property of being balanced generalizes and relaxes the definition of maximal evenness. However, a balanced rhythm may be far from maximally even. Therefore a balanced rhythm tends to have its onsets either almost evenly distributed along the circle or clustered into two groups located diametrically opposite each other.
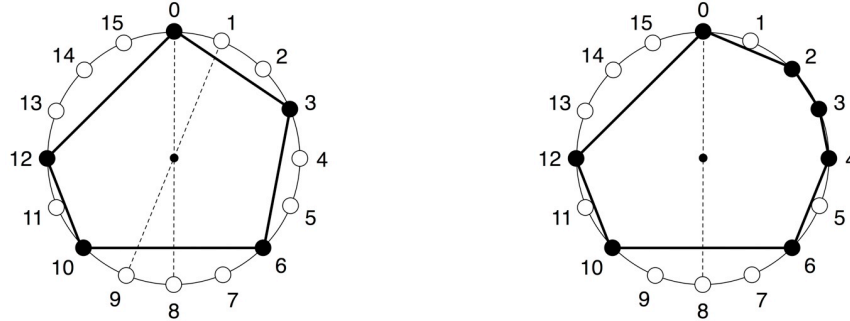

Figure 13. A balanced rhythm (left) and an unbalanced one (right).

That the fourteen features described above have proved to be useful for discourse on the theory of musical rhythm, as well as for the classification of musical rhythms in ethnomusicology, is not sufficient reason to believe that they are relevant for characterizing rhythm similarity. Indeed, for any one of these fourteen features the reader can easily construct examples of rhythms that sound different, but have the same value of the feature. The question investigated in this study is whether by using all fourteen features together, the rhythms in the resulting 14-dimensional feature space, exhibit similarity groupings that correlate with human perception.

3.   Calculating Rhythm Dissimilarity in Feature Space

Given a set of features such as the fourteen features of the cyclic rhythms described in the preceding section, any rhythm may be represented by a feature vector, in this case a 14-dimensional vector of feature values. In general, given two N-dimensional feature vectors such as $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$ obtained by calculating N feature values for each of the rhythms, the dissimilarity between the rhythms may be measured by any of a large variety of metrics or more general measures (Polansky, 1996). Perhaps the most well-known and frequently used metrics are two special cases of the Minkowski p-metrics: the Euclidean distance ($p = 2$) and the city-block distance (also Manhattan metric with $p = 1$) (Tenney & Polansky, 1980). The Euclidean distance between $\mathbf{X}$ and $\mathbf{Y}$ is given by the formula:

$$d_2(\mathbf{X},\mathbf{Y}) = (|X_1 - Y_1|^2 + |X_2 - Y_2|^2 + \ldots + |X_N - Y_N|^2)^{1/2}.$$

This distance measure imposes a particular function for combining the values of the N features to obtain a measure. In the absence of prior knowledge about the nature of

the inter-relationships that may exist between the different features, Tenney & Polansky (1980) argue in favor of using the city-block distance (also called the L_1 norm) given by the simpler sum of the absolute values of the differences of each coordinate, that is:

$$d_1(\mathbf{X}, \mathbf{Y}) = |X_1 - Y_1| + |X_2 - Y_2| + \ldots + |X_N - Y_N|.$$

This is the metric that we used in the calculation of the feature-based distance matrix. As an example, consider the cinquillo-variant and tango-congo rhythms in Figure 15. The reader may verify that the 14 preceding features calculated on these rhythms give rise to the feature vectors:

$$\mathbf{X} = \text{(cinquillo-variant)} = (1\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 2\ 2\ 0\ 0\ 0)\ \text{and}$$
$$\mathbf{Y} = \text{(tango-congo)} \quad = (1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 0\ 1\ 1).$$

Then the L_1 distance between $\mathbf{X}$ and $\mathbf{Y}$ is given by

$$d_1(\mathbf{X}, \mathbf{Y}) = (0 + 1 + 1 + 0 + 1 + 0 + 0 + 0 + 1 + 1 + 1 + 0 + 1 + 1) = 8.$$

4.   The Transformational Approach

One of the most popular distance measures used for measuring music similarity in music information retrieval applications is the edit (also Levenshtein) distance (Orpen & Huron, 1992; Smith, McNab, & Witten, 2008). Previous experiments have been reported suggesting that the edit distance correlates highly with human perceptual judgments, even in the absence of metric priming (Toussaint, Campbell, & Brown, 2010a), and that it is superior in this regard to the more constrained swap distance described in the introduction (Toussaint, Campbell, & Brown, 2010b). For these reasons the edit distance was chosen as a good representative of transformation methods to be compared to the feature based approach.

The edit distance is based on three simple mutation operations that may be performed on a rhythm, that alter the rhythm by either changing a symbol, changing the overall duration of the rhythm (longer or shorter), or both. Given two sequences of symbols, the edit distance is defined as the minimum number of substitutions, insertions, and deletions, necessary to transform one sequence to the other. The edit distance may be calculated using dynamic programming which has a computational complexity proportional to the product of the lengths (number of symbols) of the two sequences being compared. For binary sequences an insertion adds a '1' or a '0' somewhere in the sequence, making it one bit longer. A deletion removes either a '1' or a '0' somewhere in the sequence making it one bit shorter. A substitution replaces a '1' by a '0' or a '0' by a '1' leaving the length of the sequence unchanged. Each of these three operations has one unit of cost. More formally then the edit distance between two binary sequences is the smallest cost of converting one binary sequence to the other using these three operations. Since each operation has unit cost, the edit distance is simply the value that minimizes the sum of the cardinalities of the three operations. Figure 14 illustrates the result of the edit distance computation on a binary example in which the symbols are onsets and rests, represented as filled and empty squares, respectively, each of which has unit time duration. Such a representation will also be referred to here as *binary onset-rest coding*. The Figure shows how the edit distance converts the cinquillo rhythm (21212) to the habanera rhythm (3122) with only two operations,

yielding an edit distance of 2. First note that there are of course many sets of operations that convert one rhythm to the other. For example, three substitutions will do the job: change pulses 3 and 6 of rhythm 21212 from rests to onsets, and pulse 6 from onset to rest. However, the transformation may be accomplished with only 2 operations. First delete the onset at pulse 7 to convert the 8-pulse rhythm 21212 to the 7-pulse rhythm 2122. Then insert a rest between pulses 2 and 3 to convert the 7-pulse rhythm to the 8-pulse habanera rhythm 3122. Note that an insertion operation is quite powerful, and can, as this case illustrates, simultaneously translate five pulses into their correct alignment. Thus adding/deleting an onset at one position may have a much stronger effect than doing so at a different position.
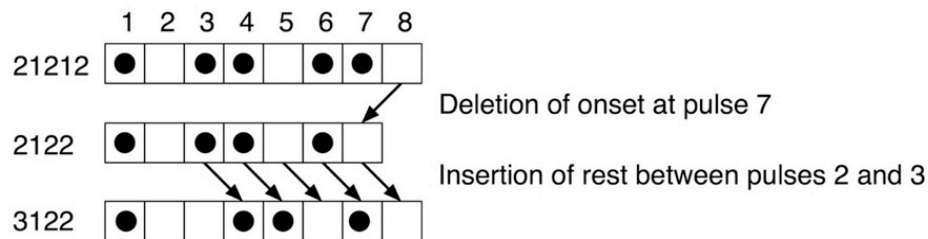

Figure 14. Calculation of the edit distance between the *cinquillo* rhythm (21212) and the *habanera* rhythm (3122).

5. Phylogenetic Analyses Methods

Distance Methods

Phylogenetic trees were originally conceived for applications in evolutionary biology for the purposes of describing and visualizing evolutionary relationships that exist between members of a group of biological organisms (Carrizo, 2004). However, more recently phylogenetic methods have been applied to a wide variety of cultural objects as well, including textiles, stone tools, variants of the Canterbury tales, and Christian denominations. (Barbrook et al. 1998; Hage et. al., 1998; Mace, Holden, & Shennan, 2005; Collard et al. 2006; Matthews 2012). Taken as a whole, these cultural phylogenetic studies have demonstrated quantitatively similar levels of decent processes as are found in biological systems (Collard et al. 2006). We see no reason *a priori* why musical study would be uniquely uninformed by phylogenetic analysis, although only most recently have such methods been applied to the domain of music (Toussaint, 2003; Toussaint, Campbell & Brown, 2010a and 2010b).

The bioinformatics and computational biology literatures are filled with a wide variety of different approaches for constructing phylogenetic trees, that fall into two main categories: distance-based methods and "character"-based methods. Distance methods assume that a distance (or dissimilarity) matrix is available containing the distance between every pair of objects being studied, in our case musical rhythms. From these distance matrices algorithms invariably construct phylogenetic trees in such a way that the minimum distance between every pair of rhythms, measured along the branches in the tree (geodesic distances), approximates as closely as

possible the corresponding distance entry in the distance matrix. In our analyses we used a popular method called *BioNJ* (Saitou and Nei, 1987; Gascuel, 1997) embedded in the software package *SplitsTree-4* (Huson, 1998; Huson & Bryant, 2006). Because the construction of phylogenetic trees in the plane necessarily involves approximating the distances in most cases, a quantitative test of the original distance matrices themselves is more accurate. The Mantel test is such a test. It belongs to the family of permutation tests for measuring the association between two distance matrices (Dietz, 1983; Hage et al., 1998), and is designed to be used in situations where the elements in the matrix are not independent, as is the case in the matrices analyzed here. We used the Mantel test software package made available by Eric Bonnet and Yves Van de Peer, with 10,000 repetitions (Bonnet & Van de Peer, 2002).

The methodology used in the study reported here is reminiscent of those previously applied in the music pitch domain by Quinn (2001), and Mavromatis and Williamson (1999a, 1999b). They applied traditional cluster analysis methods to compare a variety of measures of similarity between musical chords. Our study focuses on rhythms rather than chords, and uses phylogenetic analysis in lieu of cluster analysis.

Character Methods

Character-based methods of phylogenetic inference do not start with a pairwise distance matrix; rather, their input data are vectors of binary, multistate, or combined binary and multistate characters (binary onset-rest coding). Biologists then search through a series of statistical models that calculate the likelihood of the observed data distributions based on models that include parameters such as the rate of change between alternate states and the variation across characters in the rates of change. (Felsenstein, 2004). We applied Bayesian methods for character-based phylogenetic inference, as these methods allow for robust yet conservative inferences by conditioning across uncertainty of the parameter estimations of the underlying transition probabilities (Huelsenbeck & Ronquist, 2001). These character based methods are preferable to many distance methods in that distance methods tend to waste data in the conversion of the actual observations into pairwise distances. The methods we use do not presume that characters change in concert with any other characters in the matrix, and the methods strive to find the rate parameters and tree structure that best explain the observed data distribution. The resultant trees constitute hypotheses for which rhythms are more related to others by cultural descent. More detailed descriptions of the MCMC methodology may be found in the papers by Tierney (1994) and Mau et al. (1999).

Monte Carlo Markov Chain Inference of the Maximum Likelihood Trees

We used the Bayesian Markov-chain Monte Carlo (MCMC) based software MrBayes V 3 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) to infer the maximum likelihood trees from the two alternate data sets, one being comprised of the 14 rhythmic features and the other a binary data set that coded each pulse of a rhythm as an onset (state 1) or a rest (state 0). We considered the state labels to be arbitrary designations analogous to the typical anatomical data used in phylogenetic analyses (Lewis, 2001). For each data set, we used the harmonic mean of MCMC searches to construct Bayes factor comparisons of alternative models for how transitions would occur between states at each character (Kass & Raftery, 1995). The Bayes factor is the ratio of the a posteriori probabilities of the competing hypotheses (given the data), or equivalently, the product of the a priori probability ratio and the likelihood ratio. Unlike the edit distance, the phylogenetic analysis does not allow states to shift or split from one character to another in the binary data set that coded each rhythm as a string of onsets and rests. Rather, the phylogenetic analysis yields estimates a set of transition rates between 0 and 1 within a single character position.

The Bayesian search proceeds via a series of proposals that randomly modify one aspect of the evolutionary model. This may be an aspect of tree topology, branch length, or rate of character evolution. The software then calculates the posterior probability of this proposal, and accepts or rejects it in proportion to its probability compared to the probability of the proposal currently in the working memory. Thus, the accepted proposals form a "chain" of proposals each with somewhat different parameter values and tree shapes. One then records proposals infrequently (every 100 or 1000 or more proposals) to a saved file. The distribution of these saved proposals represents the posterior distribution of trees and parameters.

The early proposals of an MCMC are typically much less likely than later ones because they are drawn from completely random locations in parameter space. Thus, these early proposals are typically discarded as the 'burn-in' period of the MCMC analysis. We assessed the appropriate sampling frequency and burn-in by visual inspection of the chain outputs to look for autocorrelation of the likelihoods and stationarity of the likelihood values. As sampling frequency decreases, one must run more 'generations' of an MCMC in order to adequately sample parameter space. We also examined the potential scale reduction factor (PSRF), a heuristic index that approaches 1 as multiple chains converge on the same parameter space.

For each data set, we ran three analyses for different models of character evolution. One model allowed for a single transition rate among the alternative states at each character. This rate was symmetric (i.e. 0 $\rightarrow$ 1 changes equaled 1 $\rightarrow$ 0 changes) and identical across characters. The other two models added parameters to relax these constraints on character change. One model, allowed for asymmetric transitions between character states according to a beta distribution (Lewis, 2001), while the other used a gamma distribution to allow for variation across the different characters such that some could evolve more quickly than others.

We compared the trees output by the best character models for each data set to the trees generated by the *BioNJ* algorithm applied to the matrices of the edit, feature, and listening experiment distances. We quantified these comparisons by calculating the partition metric, which is a standard measure of tree similarity that corresponds to twice the number of differing bipartitions specified by internal branches of a phylogeny (Penny and Hendy 1985). We assessed significance of the partition metrics by comparing the observed values to the null distribution obtained by generating 1000 random tree topologies. These analyses were performed with the R statistical software and the "ape" package.

6. The Rhythm Data

The data used in the experiments consisted of the nine rhythms, expressed in circular and polygon notation in Figure 15, taken from Mario Rey's ethnographic study of Cuban art music (Rey, 2006). Mario Rey classified the seven rhythms in Figure 15 (other than 2-3-3 and 3-2-3), which represent some of the most frequently used Afro-Cuban rhythms, into two groups: the *cinquillo* and *cinquillo-variant* derived from the *contradanza*, and the *tresillo*, *tango-congo* and *conga* derived from the *habanera*. The *tresillo* may be derived from the *habanera* by substituting the third onset for a rest, the *tango-congo* by substituting the last onset for a rest, and the *conga* by substituting the two last onsets for rests. On the other hand the *cinquillo* may be derived from the *contradanza* by substituting the fourth onset of the *contradanza* for a rest, and the *cinquillo-variant* may be obtained by substituting the fourth onset for a rest, and

substituting the rest at pulse 7 for an onset. Of course, these transformations are not necessarily those made by the edit distance calculations. Rey also suggested that the *habanera* and *contradanza* rhythms were the ancestral rhythms in their respective groups. The rhythms 2-3-3 and 3-2-3 were added to the data out of curiosity because they are rotations of the *tresillo* and are used in traditional music all over the world.

In a previous investigation the edit distance confirmed the two-group categorization of Rey's ethnographic study, and provided evidence that the *habanera* serves as a prototype for this collection of rhythms, thus supporting Rey's assertion that one group derives from the *habanera* (Toussaint, Campbell, & Brown, 2010a). However, no support was found for the claim that the *contradanza* played an ancestral role.
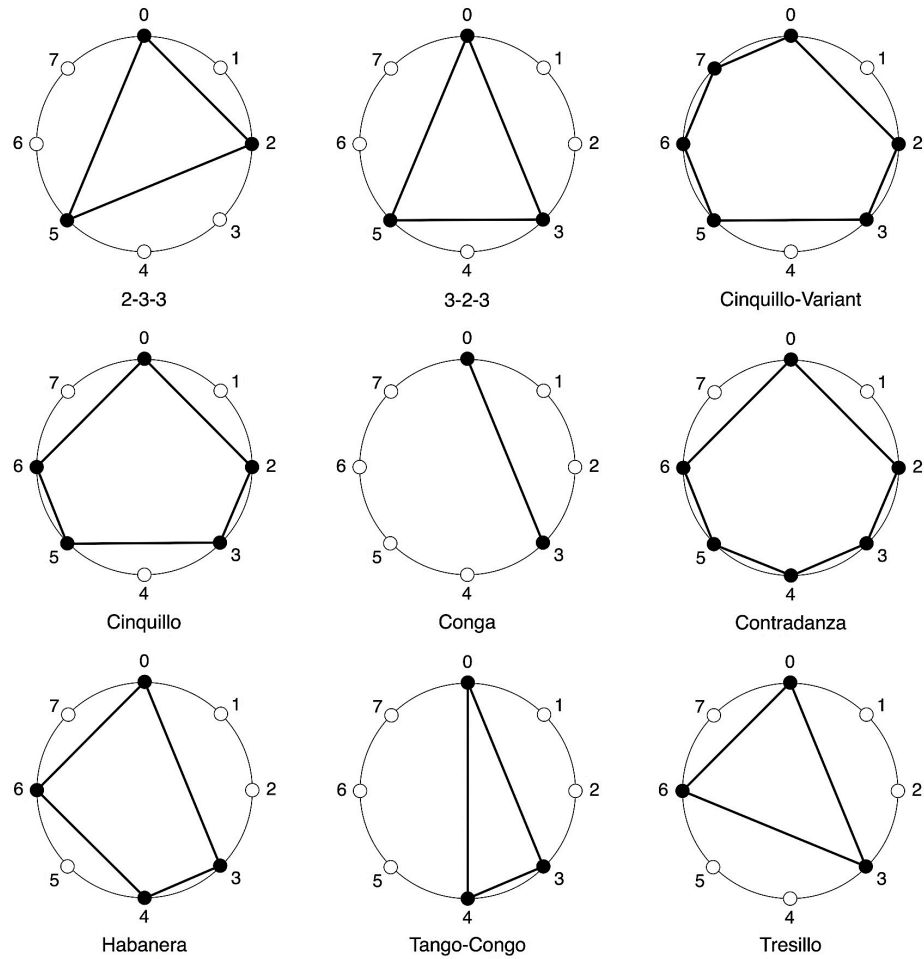
Figure 15. The nine rhythms used in the experiments.

7. Listening Experiments

For completeness, the details of the listening experiments reported previously (Toussaint, Campbell, and Brown, 2010a), showing that the edit distance is a good predictor on human perceptual judgments, are briefly described below.

Participants

Sixteen participants (8 males and 8 females) took part in the listening tests, and ranged in age from 18 to 57 (mean age = 29.5). Eight participants were 2009-2010 Radcliffe Fellows at the Radcliffe Institute for Advanced Study at Harvard University, and eight were undergraduate students at Harvard University majoring in music. Each participant was paid $10 for his/her participation. The Radcliffe Fellows represented a variety of academic disciplines. The sixteen participants had an average of 9.1 years of musical training. Some of the Radcliffe Fellows had no musical training.

Apparatus

The participants listened to the rhythms using either Sennheiser, model PXC 250, or SONY, model MDR-NC7, noise-cancelling headphones (*NoiseGuard^{TM}*). These headphones were connected to a MacBook Pro laptop Apple computer which displayed the graphical user interface of the *Sonic Mapper* software developed by Gary P. Scavone using *Qt* for the user interface and *RtAudio* for audio output (Scavone, Lakatos, & Harbke, 2002). The *Sonic Mapper* software offers a variety of experimental methods for comparing acoustic inputs. In these experiments the pairwise comparison method was used.

Stimulus materials

In order to create sound samples that resembled the binary symbolic notated rhythms as closely as possible, the sound samples (created using Apple *Garageband*), were entered exactly in MIDI format. Each onset triggered an identical clinical click, which resembled the sound of a pair of wooden claves. Each rhythm was repeated four times in succession at a tempo of 200 pulses per minute, resulting in a sound sample that lasted for 8 seconds.

Procedure

Before the start of the experiment each participant filled out a form with some biographical data, and read some instructions that told the participants that they would be hearing 36 pairs of rhythms, and that they would be asked to compare them using a pseudo-continuous sliding scale from 1 (most dissimilar on the left) to 10 (most similar on the right). The *SonicMapper* program presented each pair of rhythms only once in a randomized order. Furthermore, all the rhythm pairs were also presented in different random orders to each participant. The entire listening test took between 20 and 30 minutes. The participants were not trained to judge the range of variability of the rhythms, and they were not primed with an underlying meter.

The scores obtained from each participant were subsequently subtracted from ten in order to convert them to dissimilarities rather than similarities, a requirement of the input format for the phylogenetic analysis software package used. Finally, for each pair of rhythms compared, the median distance across all the participants was used in the Phylogenetic analyses and Mantel tests.

8. Results

Results from the Distance-based Method

The feature vectors obtained from the fourteen features computed for each of the nine rhythms of Figure 15 are shown in Figure 16.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2-3-3	1	0	0	0	2	0	0	1	0	0	1	0	1	1
3-2-3	1	0	0	0	2	1	0	0	0	0	1	0	1	1
Cinquillo-Variant	1	1	1	0	1	0	0	0	1	2	2	0	0	0
Cinquillo	1	0	1	0	2	1	0	0	0	1	1	0	1	1
Conga	1	1	0	0	1	0	0	1	0	0	1	0	1	0
Contradanza	1	1	1	0	1	1	0	0	0	2	1	1	0	0
Habanera	1	1	0	0	1	0	0	0	0	1	1	0	1	0
Tango-Congo	1	0	0	0	0	0	0	0	0	1	1	0	1	1
Tresillo	1	0	0	0	2	0	0	0	1	0	1	0	1	1

Figure 16. The 14-dimensional feature vectors for each of the nine rhythms of Figure 15.

The dissimilarity matrices obtained from the listening experiments with human subjects, the L_1 metric (city-block distance) in 14-dimensional feature space, and the edit distance, are shown in Figures 17, 18, and 19, respectively. The bottom rows in the tables, labeled TOTAL, list for each rhythm the sum of its dissimilarities to all the other rhythms. The corresponding phylogenetic trees for these three matrices appear in Figures 20, 21, and 22, respectively.

In Figure 17 the rhythm most similar to all the others (one of the definitions of a prototype) is the *tresillo*, with a lowest TOTAL score of 30.72. The *habanera* on the other hand, with a score of 34.36, comes in second as a candidate for prototype. The term prototype is used here to refer to a *good* exemplar or a highly representative instance of a category as is done by Rosch (1975) and MacLaury (1991).

The phylogenetic tree in Figure 20 computed from the matrix obtained from the human similarity judgments of Figure 17 separates the *contradanza*, *cinquillo*, and *cinquillo-variant* group from the remaining rhythms, in agreement with Rey's classification (Rey, 2006). A detailed description of the listening tests and their results appears in Toussaint, Campbell, & Brown, (2011). The results varied

considerably among subjects, with standard deviations ranging from 1.35 to 2.26, with an average of 1.7.

Concerning the L_1 metric (city-block distance) matrix computed with the 14 features, there are some marked differences between the tree from the listening experiments (Figure 20) and that generated by the 14 rhythmic features (Figure 21). Furthermore, the Mantel test performed on the corresponding two matrices used to infer the trees gave a very low correlation coefficient of $r = 0.181$ with a p-value of 0.138 (one-tailed test). In spite of these overall poor results, the feature-based method yielded one local agreement with the listening results: the *tresillo* and rhythm 2-3-3 are clustered together on both trees. Furthermore, the feature-based method yields the *habanera* as the prototype, with a TOTAL score of 35.

The tree in Figure 22 obtained from the edit distance matrix, like the tree resulting from the listening tests, separates the *contradanza*, *cinquillo*, and *cinquillo-variant* group from the remaining rhythms, in agreement with Rey's classification (Rey, 2006), and exhibits several additional local similarities with the tree shown in Figure 20. Compared to the result for the features, the Mantel test calculated on the edit distances and the scores from the listening experiments gave a much higher correlation coefficient of $r = 0.594$ with a p value of 0.0002 (one-tailed test). The edit distance also yields the *habanera* as a prime candidate for the best prototype, with a TOTAL score of 16 (tied with the *tresillo* and *cinquillo*).

Results from MCMC Inference of the Maximum Likelihood Trees

The onset-rest and feature trees with the highest posterior probabilities are shown in Figures 23 and 24, respectively. Because we considered all tree topologies to have equal prior probabilities (before the analysis), these trees with the highest posterior probabilities are also estimates of the maximum likelihood trees for the data. For the feature data set, visual inspection of the chain likelihoods and the PSRF diagnostic (potential scale reduction factor) lead us to use six independent runs each of 5,000,000 generations, sampled every 1000 generations. The chains from the binary onset/rest coding data appear less autocorrelated, so we used 100,000 generations with a sampling frequency of 100. In both cases we used burn-in values of 100, thereby discarding the first 100 saved proposals from the posterior draws.

The maximum likelihood tree from the binary onset/rest coded rhythms shown in Figure 23, agrees with the edit-distance tree (Figure 22) and the listening test tree (Figure 20), in separating the *contradanza*, *cinquillo*, and *cinquillo-variant* group from the remaining rhythms, again in agreement with Rey's classification (Rey, 2006). It also agrees in clustering the *habanera* and *tango-congo* together. Indeed, this tree was most similar to the *BioNJ* tree inferred from the edit distances (partition metric = 6). This similarity was significant as only 4 of 1000 random tree topologies had partition metrics equal to or less than 6 ($p = 0.004$). The binary rhythm tree was also significantly similar to the tree inferred from the listening experiments (partition metric = 8, $p = 0.036$). However, the binary rhythm tree was not significantly similar to the tree inferred from the feature-based distances (partition metric = 10, $p = 0.221$)

The maximum likelihood tree from the features of the rhythms was not similar to either the edit distance tree or to the tree from listening experiments (partition metric = 14, $p = 1.000$). In fact, the feature tree was as different from the edit distance and listening experiment trees as it was from the most different of the random trees in our null distribution. The maximum likelihood feature tree was similar, however, to the tree inferred from the matrix of feature distances (partition metric = 6, $p < 0.001$).

	2-3-3	3-2-3	cinquillo -variant	cinquillo	conga	contradanza	habanera	tango- congo	tresillo
2-3-3	0	4.52	5.16	4.64	5.8	4.64	5.76	5.8	2.36
3-2-3	4.52	0	5.12	4.68	3.32	4.72	4.08	5.4	3.6
cinquillo-variant	5.16	5.12	0	2.52	4.44	3.72	4.6	5.28	3.56
cinquillo	4.64	4.68	2.52	0	5.6	4.28	5.04	5.32	3.96
conga	5.8	3.32	4.44	5.6	0	5.92	4.48	4.24	2.28
contradanza	4.64	4.72	3.72	4.28	5.92	0	4.24	4.44	5
habanera	5.76	4.08	4.6	5.04	4.48	4.24	0	2.24	3.92
tango-congo	5.8	5.4	5.28	5.32	4.24	4.44	2.24	0	6.04
tresillo	2.36	3.6	3.56	3.96	2.28	5.0	3.92	6.04	0
TOTAL	38.68	35.44	34.4	36.04	36.08	36.96	34.36	38.76	30.72

Figure 17. The dissimilarity matrix obtained from the listening experiments.

	2-3-3	3-2-3	cinquillo -variant	cinquillo	conga	contradanza	habanera	tango- congo	tresillo
2-3-3	0	2	10	4	3	10	5	4	2
3-2-3	2	0	10	2	5	8	5	4	2
cinquillo-variant	10	10	0	8	7	4	5	8	8
cinquillo	4	2	8	0	7	6	5	4	4
conga	3	5	7	7	0	7	2	5	5
contradanza	10	8	4	6	7	0	5	8	10
habanera	5	5	5	5	2	5	0	3	5
tango-congo	4	4	8	4	5	8	3	0	4
tresillo	2	2	8	4	5	10	5	4	0
TOTAL	40	38	60	40	41	58	35	40	40

Figure 18. The L_1 distance matrix computed with the 14 features.

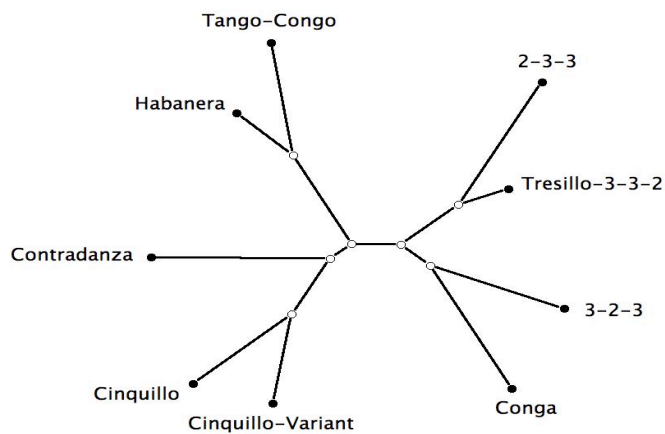|	2-3-3	3-2-3	cinquillo -variant	cinquillo	conga	contradanza	habanera	tango- congo	tresillo
2-3-3	0	2	3	2	2	3	3	3	2
3-2-3	2	0	3	2	1	3	2	2	2
cinquillo-variant	3	3	0	1	4	2	3	4	3
cinquillo	2	2	1	0	3	1	2	3	2
conga	2	1	4	3	0	4	2	1	1
contradanza	3	3	2	1	4	0	2	3	3
habanera	3	2	3	2	2	2	0	1	1
tango-congo	3	2	4	3	1	3	1	0	2
tresillo	2	2	3	2	1	3	1	2	0
TOTAL	20	17	23	16	18	21	16	19	16

Figure 19. The edit distance matrix.

Figure 20. The phylogenetic tree computed from the dissimilarity matrix obtained from the human similarity judgments.
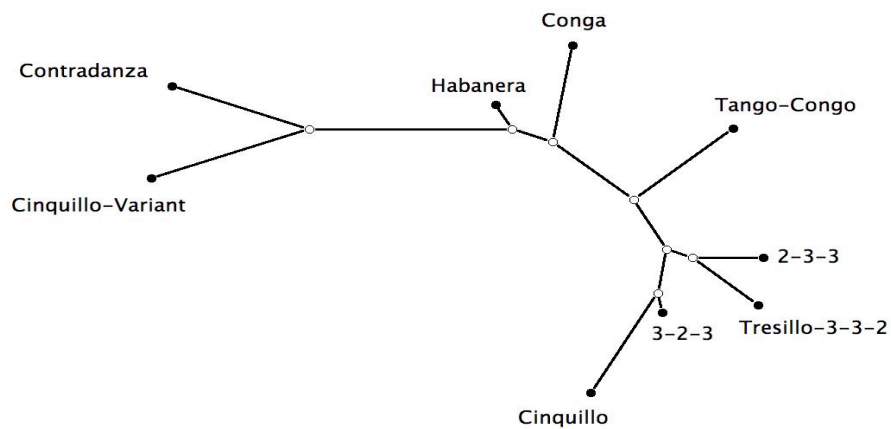

Figure 21. The phylogenetic tree computed from the L_1-distance matrix using the 14-dimensional feature vectors.
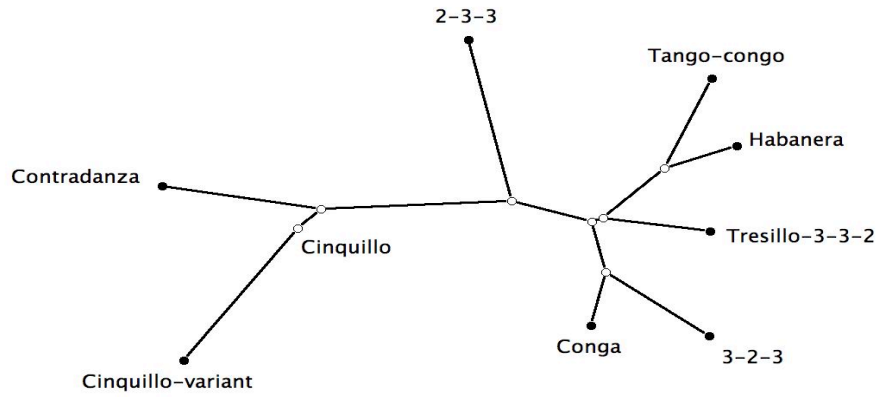
Figure 22. The phylogenetic tree computed from the edit-distance matrix.
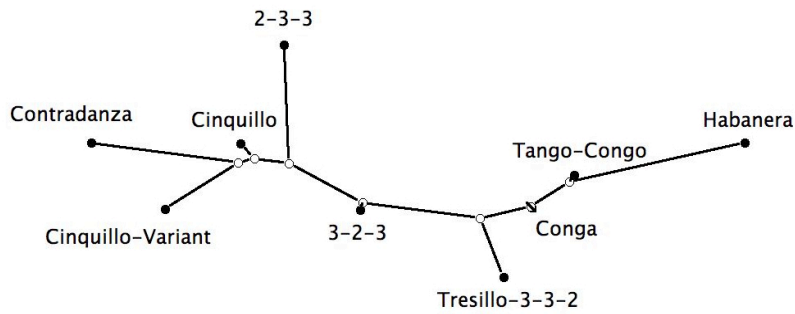
Figure 23. The maximum posterior probability trees for the binary onset-rest coding of the rhythms.
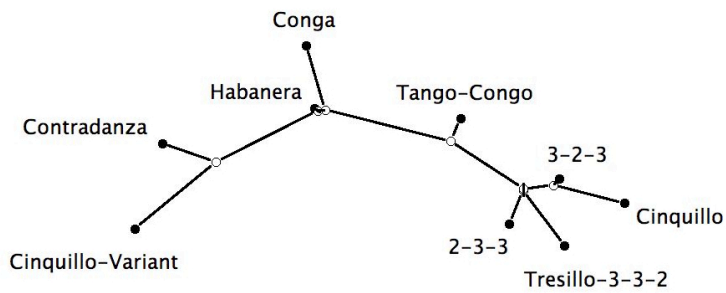
Figure 24. The maximum posterior probability trees for the features of the rhythms.

9. Discussion

Two very different implementations of comparing rhythms based on transformations (*BioNJ* and Bayesian character evolution) are much better at recovering human judgments and Rey's ethnographic account (Rey, 2006) than are the same algorithms applied to comparisons of rhythms based on features. What the edit distance and the Bayesian analysis of the binary onset/rest codes have in common is that they both model the transformation of one rhythm into exactly another rhythm. In this sense they are both 'transformational' even though they imply different underlying mechanisms of how these transformations occur. Their congruences with the listening experiments and with Rey's ethno-musicological account suggest that transformational similarity and difference (conversion of one sequence into another) are most salient to individuals and to cultural processes.

The lack of congruence of the feature-based methods (either *BioNJ* or Bayesian) with the listening experiments and Rey's ethnographic account suggests that features of rhythms are of little importance to human cognition. The listening experiments indicate that people are either not attentive to mathematical features such as evenness or symmetry, or else they are incapable of recognizing when such features are similar or different. To the extent that Rey's ethnographic account of culturally related rhythms is accurate, it makes sense that features would also be unimportant in rhythmic evolution since the learners of the rhythms either pay no attention to or cannot process the features.

In spite of the overall lack of congruence of either of the two feature-based methods (*BioNJ* or Bayesian) with Rey's ethnographic account, the feature-based distance method using the L_1-distance does manage to single out the *habanera* as the rhythm from which *all* the others are derived, with a TOTAL score of 35 (Figure 18), thus adding a twist to Rey's account.

Although the results provide significant insights into the processes by which human judgments and rhythmic evolution may occur, the transformational approach provides limited ability to reconstruct deep ethno-historical lineages of rhythms. We found very low statistical support for the tree topologies depicted, and phylogeneticists would regard these trees as undifferentiated polytomies due to this low support. This does not invalidate the present findings because our focus is on how the optimal trees compare to human similarity judgments and cultural processes that are known from data sets that are completely independent of the inference of the phylogenetics trees (that is, our listening experiments and Rey's ethonographic account). The question of the appropriateness of the application of phylogenetic tools from biology to other domains such as cultural evolution in general and musical rhythm in particular, has already been discussed in the literature. While it is true that in this paper the trees are used mainly for the purpose of visualizing the structural relationships between the rhythms, in previous research we have found evidence in similar applications of phylogenetic trees to rhythms from other cultures such as flamenco meters (Diaz-Bañez, et. al., 2004), North Indian talas (Thul & Toussaint, 2008b), and Arabic rhythms (Toussaint, Campbell, & Brown, 2011), that the central prototypes do play an ancestral role in the evolution of those rhythms. Other researchers have also found that the phylogenetic tools used in biology are applicable to cultural evolution as well. See for example the paper on the evolution of carpet designs by Collard, & Tehrani (2005), and the study of cultural inheritance versus diffusion of religious violence (Matthews, Edmonds, Wildman, & Nunn, 2012).

Additionally, the incongruence of feature-based approaches means that rhythm phylogenetic studies will have difficulty extending beyond comparisons where

different rhythms can be placed (e.g. by expansion or contraction of the timespan of the rhythmic cycle) on the same scale. When one rhythm is not evenly divisibly by another, it is much more difficult to line up the binary strings of onsets and rests in a preliminary hypothesis of which character states are transitioning to others. This is required of transformational techniques. Features, on the other hand, can be calculated on and compared for any set of rhythms that differ in both their number of onsets as well as their number of pulses, but our results suggest that features of rhythms, play no role in human rhythmic judgments or cultural processes. However, these conclusions should be taken with a grain of salt, since they are based on only one family of nine rhythms. Further experiments with other data sets are required to make firmer conclusions on this issue.

It is also worth emphasizing that the conclusions offered here may apply only to *structural* features such as those considered in this study. For such features there is a clear distinction between *feature*-based and *transformation*-based measures of similarity. By contrast, for descriptions of rhythms based on simpler local features, the distinction between the two approaches may become blurred. A concrete rather extreme example will clarify this point. Consider a rhythm in binary sequence notation such as the clave son [x . . x . . x . . . x . x . . .] consisting of sixteen pulses with five onsets and eleven rests. This rhythm may be viewed as a point in the 5-dimensional onset-feature-space, consisting of five local duration features. Each feature is an onset characterized by the time elapsed between the start of the rhythm and the time the onset is heard. In other words, the feature vector of this rhythm is given by [0, 3, 6, 10, 12]. Compare this rhythm to another rhythm [x . x . . . . . x . x . . . x .] with feature vector [0, 2, 8, 10, 14]. Computing the L_1-distance between these two feature vectors yields the calculation $(0 + 1 + 2 + 0 + 2) = 5$. The transformation method described in the introduction that transforms one rhythm to another by means of the minimum number of *swaps* between adjacent onsets and rests yields a swap distance also equal to 5. Clearly, although this feature-based method appears superficially to be different from this transformation method, the two are merely alternate *implementations* of the same underlying measure, and thus in this case both methods will yield the same results.

Finally it is worth pointing out that the features selected in this study were not chosen on the basis of their lack of correlation, but because they were important in music theory. It would be interesting to determine how correlated these theoretical features are for rhythms used in practice. Some features, such as the toggle and off-beatness properties are probably correlated. Better results could conceivably be obtained using a totally different set of uncorrelated features. However, such an investigation was beyond the scope of this study, and is left for future research.

10.   Acknowledgment

References

Agawu, K. (2006). Structural analysis or cultural analysis? Competing perspectives on the *standard pattern* of West African rhythm. *Journal of the American Musicological Society, 59*(*1*), 1-46.

Amiot, E. (2007). David Lewin and maximally evens sets. *Journal of Mathematics and Music*, *2*, 157-172.

Arom, S. (2004). *African polyphony and polyrhythm,* Cambridge University Press.

Arom, S. (2001). Prolegomena to a biomusicology. in *The Origins of Music*, Eds., Nils L. Wallin, Bjorn Merker, and Steven Brown, The MIT Press, Cambridge, MA, 27-29.

Barbrook, A. C., Howe, C. J., Blake, N., & Robinson, P. (1998). The phylogeny of *The Canterbury Tales*. *Nature*, *394*, p. 839.

Berenzweig, A., Logan, B., Ellis, D. P. W., & Whitman, B. (2004). A large-scale evaluation of acoustic and subjective music-similarity measures. *Computer Music Journal*, *28*(*2*), Music Information Retrieval, 63-76.

Bonnet, E., & Van de Peer, Y. (2002). zt: a software tool for simple and partial Mantel tests. *Journal of Statistical Software*, *7*(*10*), 1-12.

Brandel, R. (1959). The African hemiola style. *Ethnomusicology*, *3*(*3*), 106-117.

Brown, A. R., Towsey, M. W., Wright, S. K., & Deiderich, J., Statistical analysis of the features of diatonic music using jMusic software. In *Proceedings Computing ARTS 2001 – Digital resources for research in the humanities*, Sydney, Australia: The University of Sydney, 1-11.

Carrizo, S. F. (2004). Phylogenetic trees: an information visualization perspective. *Proceedings of the Second Conference on Asia-Pacific Bioinformatics*, *29*, 315-320.

Cerulo, K. A. (1988). Analyzing cultural products: A new method of measurement. *Social Science Research*, *17*, 317-352.

Chemlillier M., & Truchet, C. (2003). Computation of words satisfying the "rhythmic oddity property" (after Simha Arom's work). *Information Processing Letters*, *86*, 255–261.

Clough J., & Douthett, J. (1991). Maximally even sets. *Journal of Music Theory*, *35*(*1/2*), Spring-Autumn, 93-173.

Collard, M., Shennan, S. J., & Tehrani, J. J. (2006). Branching, blending, and the evolution of cultural similarities and differences among human populations. *Evolution and Human Behavior*, *27*, 169–184.

Collard, M. & Tehrani, J. (2005) Cladistic analysis of Turkmen textiles sheds light on cultural evolution. In: Mace, R. and Holden, C.J. and Shennan, S., (eds.) *The evolution of cultural diversity: a phylogenetic approach*. (pp. 109-131). Left Coast Press: Walnut Creek, US.

Crawford, T., Iliopoulos, C. S., & Raman, R. (1997-98). String-matching techniques for musical similarity and melodic recognition. *Computing in Musicology*, *11*, 73-100.

Dean, R. T., Byron, T., & Bailes, F. A. (2009-10). The pulse of symmetry: On the possible co-evolution of rhythm in music and dance. *Musicæ Scientiæ*, Special Issue on Music and Evolution, 341-367.

Demorest, S. M., and Serlin, R. C. (1997). The integration of pitch and rhythm in musical judgment: Testing age-related trends in novice listeners. *Journal of Research in Music Education*, *45*(*1*), 67-79.

Diaz-Bañez, M., Farigu, G., Gomez, F., Rappaport, D. and Toussaint, G. T. (2004). El compás flamenco: A phylogenetic analysis. *Proceedings of BRIDGES: Mathematical Connections in Art, Music, and Science,* Southwestern College, Winfield, Kansas, July 30 to August 1, pp. 61-70.

Dietz, E. J. (1983). Permutation tests for association between two distance matrices. *Systematic Zoology*, *32*(*1*), 21-26.

Domínguez, M., Clampitt, D., & Noll, T. (2009). WF scales, ME sets, and Christoffel words. *Mathematics and Computation in Music*, T. Klouche and T. Noll, Eds., Springer-Verlag, Berlin, Heidelberg, 477-488.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification*, Wiley-Interscience, 2nd Edition.

Eerola, T., Järvinen, T., Louhivuori, J., & Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Perception*, *18*(*3*), 275-296.

Feldman, M. (1981). Crippled symmetry. *Anthropology and Aesthetics*, *2*, 91-103.

Felsenstein, J. (2004). *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates, Inc.

Ferraro, P., & Hanna, P. (2007). Optimizations of local edition for evaluating similarity between monophonic musical sequences. *Proceedings of the 8th International Conference on Information Retrieval*, Pittsburgh PA, U.S.A. May 30-June 1, 2007.

Fracile, N. (2003). The *Aksak* rhythm, a distinctive feature of the Balkan folklore. *Studia Musicologica Academiae Scientiarum Hungaricae*, *44*, 197-210.

Guastavino, C., Gómez, F., Toussaint, G. T., Marandola, F., & Gómez, E. (2009). Measuring similarity between flamenco rhythmic patterns. *Journal of New Music Research, 38*(2), 129–138.

Gusfield, D. (1997). *Algorithms on strings, trees, and sequences: Computer science and computational biology*, Cambridge University Press, Cambridge.

Hage, P., Harary, F., & Krackhardt, D. (1998). A test of communication and cultural similarity in polyhesian prehistory. *Current Anthropology*, *39*(5), 699-703.

Hahn, U., Chater, N., & L. B. Richardson, L. B. (2003). Similarity as transformation. *Cognition*, *87*, 1-32.

Hahn, U., & Chater, N. (1998a). Similarity and rules. Distinct? Exhaustive? Empirically distinguishable? *Cognition*, *65*, 197–230.

Hahn, U., & Chater, N. (1998b). Understanding similarity: a joint project for psychology, case-based reasoning and law. *Artificial Intelligence Review*, *12*, 393–427.

Hahn, U., & Chater, N. (1997). Concepts and similarity. In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts and Categories*, Hove: Psychology Press, 43-92.

Herrero, O. F. (2910). Symmetry in music. *Journal of Physics: Conference Series 237*, Symmetries in Science XIV, 1-11.

Hodges, W. (2003). The geometry of music. *Music and Mathematics: From Pythagoras to Fractals*, Eds., J. Fauvel, R. Flood, and R. Wilson, Oxford University Press, Oxford, 91-111.

Hodgetts, C. J., Hahn, U., & Chater, N. (2009). Transformation and alignment in similarity. *Cognition*, *113*, 62-79.

Howard, K. (1992). Why do it that way? Rhythmic models and motifs in Korean percussion bands. *Asian Music*, *23*(1), 1-59.

Huelsenbeck, J. P., & Ronquist, F. (2001). MrBayes: Bayesian inference of phylogeny. *Bioinformatics*, *17*, 754-755.

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution, 23*(2), 254-267.

Huson, D. H. (1998). *SplitsTree*: A program for analyzing and visualizing evolutionary data. *Bioinformatics*, *14*(10), 68–73.

Jiang, M. (2008). On the sum of distances along a circle. *Discrete Mathematics*, 308, 2038-2045.

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association, 90*(430), 773-795.

Kempf, D. (1996). What is symmetry in music? *International Review of the Aesthetics and Sociology of Music*, *27*(2), 155-165.

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, *50*(6), 913-925.

Lezcano, J. M. (1991). African-derived rhythmical and metrical elements in selected songs of Alejandro García Caturla and Amadeo Roldán. *Latin American Music Review*, *12*(2), 173-186.

Liebermann, P., & Liebermann, R. (1990). Symmetry in question and answer sequences in music. *Computers and Mathematics with Applications*, *19*(7), 59-66.

London, J. (2009). The effect of structural and performance factors in the perception of anacruses. *Music Perception*, *27*(2), 103-120.

Lowrance, R., & Wagner, R. A. (1975). An extension of the string-to-string correction problem. *Journal of the Association for Computing Machinery*. *22*, 177-183.

Mace, R., Holden, C. J., & Shennan, S. (Eds.) (2005), *The evolution of cultural diversity: A phylogenetics approach*, London: UCL Press.

MacLaury, R. E. (1991). Prototypes revisited. *Annual Review of Anthropology, 20*, 55–74.

Magill, J. M., & Pressing, J. L. (1997). Asymmetric cognitive clock structures in West African rhythms. *Music Perception: An Interdisciplinary Journal*, *15*(2), 189-221.

Matthews, L.J. (2012). The recognition signal hypothesis for the adaptive evolution of religion: a phylogenetic test with Christian denominations. *Human Nature, 23*, 218-

249.

Matthews, L. J., Edmonds, J., Wildman, W. J., & Nunn, C. L. (2012). Cultural inheritance or cultural diffusion of religious violence? A quantitative case study of the Radical Reformation, *Religion, Brain & Behavior*, 2:1-13.

Mau, R., Newton, M. A., & Larget. B. (1999). Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics*, *55*, 1-12.

Mavromatis, P., & Williamson, V. (1999a). Toward a perceptual model for categorizing atonal sonorities. *The Annual Conference of the Society for Music Theory*, Atlanta, Georgia, November 1999.

Mavromatis, P., & Williamson, V. (1999b). Categorizing atonal sonorities: Multidimensional scaling, tree-fitting and clustering compared. *The Annual Conference of the Society for Music Perception and Cognition*, Evanston, Illinois, August 1999.

Nilsson, N. J. (1990). *The mathematical foundations of learning machines*, Morgan Kaufmann Publishers, San Mateo, California.

Olson, R. K., & Attneave, F. (1970). What variables produce similarity grouping? *The American Journal of Psychology*, *83*(1), 1-21.

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

Penny, D., & Hendy, M. D. (1985). The Use of Tree Comparison Metrics. *Systematic Zoology*, *34*(1), 75-82.

Polansky, L. (1996). Morphological metrics. *Journal of New Music Research*, *25*, 289-368.

Porter, D. H. (1970). Reflective symmetry in music and literature. *Perspectives of New Music*, *8*(2), 118-122.

Proca-Ciortera, V. (1969). On rhythm in Rumanian folk dance. *Yearbook of the International Folk Music Council*, *1*,176-199.

Quinn, I. (2001). Listening to similarity relations. *Perspectives of New Music*, *39*(2), 108-158.

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

Rahn, J. (1987). Asymmetrical ostinatos in Sub-Saharan music: time, pitch, and cycles reconsidered. *In Theory Only: Journal of the Michigan Music Theory Society*, *9*(*7*), 23– 37.

Rey, M. (2006). The rhythmic component of *Afrocubanismo* in the art music of Cuba. *Black Music Research Journal*, *26*(*2*), 181-212.

Rivière, H. (1993). On rhythmical marking in music. *Ethnomusicology*, 37(2), 243-250.

Rosch, E. (1975). Universals and cultural specifics in human categorization. In *Cross-Cultural Perspectives on Learning*, edited by R. Breslin, S. Bochner, & W. Lonner. New York: Halsted Pres

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*, 1572-1574.

Saitou, N., & Nei, M. (1987). The Neighbor-Joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*, 406–425.

Sankoff, D., & Kruskal, J. (Eds.). (1999). *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison*, CLSI Publications, Stanford University, 1999.

Shepard, R. N. (1957). Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325–345.

Smith, L. A., McNab, R. J., & Witten, I. H. (2008). Sequence-based melodic comparison: A dynamic programming approach. *Computing in Musicology*, *11*, 101-117.

Temperley, D. (2010). Modeling common-practice rhythm. *Music Perception*, *27*(5), 355-376.

Tenney, J. & Polansky, L. (1980). Temporal Gestalt perception in music. *Journal of Music Theory*, *24*(2), 205-241.

Thul, E. & Toussaint, G. T. (2008a). Rhythm complexity measures: A comparison of mathematical models of human perception and performance. In *Proceedings of the 9th International Society for Music Information Retrieval Conference*, (ISMIR 2008), pages 663--668, Philadelphia, USA.

Thul, E. & Toussaint, G. T., (2008b). A comparative phylogenetic analysis of African *timelines* and North Indian *talas*. The 11th Annual Bridges Conference, *BRIDGES LEEUWARDEN: Mathematics, Music, Art, Architecture, Culture*, July 24-28, pp. 187-191.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of*

Statistics, 22(4):1701-1762.

Toussaint, G. T. (2011). The rhythm that conquered the world: What makes a "good" rhythm good. *Percussive Notes*, November 2011, pp. 52-59.

Toussaint, G. T. (2010). Generating "good" musical rhythms algorithmically. *Proceedings of the 8th International Conference on Arts and Humanities*, Honolulu, Hawaii, January 13-16, 774-791.

Toussaint, G. T., Campbell, M., & Brown, N. (2011). Computational models of symbolic rhythm similarity: Correlation with human judgments. *Analytical Approaches to World Music*, *1(2)*.

Toussaint, G. T., Campbell, M., & Brown, N. (2010a). Comparing mathematical measures of rhythm similarity with human perception. *Abstracts of the Third Meeting of the Northeast Music Cognition Group*, (NEMCOG), Berklee College of Music, Boston, MA.

Toussaint, G. T., Campbell, M., & Brown, N. (2010b). Modeling rhythm similarity: Experiments with Afro-Cuban timelines. *Abstracts of the Second Meeting of the Northeast Music Cognition Group*, (NEMCOG), Department of Music, Yale University, New Haven, CT.

Toussaint, G. T. (2005a). The geometry of musical rhythm. *Proceedings of the Japan Conference on Discrete and Computational Geometry*, J. Akiyama et al. (Eds.), LNCS 3742, Springer-Verlag, Berlin, Heidelberg, 198-212.

Toussaint, G. T. (2005b). The Euclidean algorithm generates traditional musical rhythms. *Proceedings of BRIDGES: Mathematical Connections in Art, Music, and Science,* Banff, Alberta, Canada, July 31 to August 3, 47-56.

Toussaint, G. T. (2005c). Mathematical features for recognizing preference in Sub-Saharan African traditional rhythm timelines. *Proceedings of the 3rd International Conference on Advances in Pattern Recognition*, University of Bath, Bath, United Kingdom, August 22-25, 18-27.

Toussaint, G. T., (2004). A comparison of rhythmic similarity measures. In *Proc. 5th International Conference on Music Information Retrieval*, Barcelona, Spain, 242-245.

Toussaint, G. T. (2003). Classification and phylogenetic analysis of African ternary rhythm timelines. *Proceedings of BRIDGES: Mathematical Connections in Art, Music, and Science,* University of Granada, Granada, Spain, 25-36.

Toussaint, G. T. (2002). A mathematical analysis of African, Brazilian, and Cuban clave rhythms. *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science,* Townson University, Towson, MD, 157-168.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327-352.

Typke, R., R. Veltkamp, R. C., and Wiering, F. (2004). Searching notated polyphonic music using transportation distances. *Proceedings of the ACM Multimedia Conference*, New-York, USA, 128-135.

Van Den Broek, E. M. F., & Todd, P. M. (2009-10). Evolution of rhythm as an indicator of mate quality. *Musicæ Scientiæ*, Special Issue on *Music and Evolution*, 69-386.

Wieling, M. B. (2007). *Comparison of Dutch Dialects*, Masters Thesis, University of Groningen, The Netherlands.

Wilson, D. (1986). Symmetry and its 'love-hate' role in music. *Computers and Mathematics with Applications*, *12B(1/2)*, 101-112.

Witten, I. H., Manzara, L. C., & Conklin, D. (1994). Comparing human and computational models of music prediction. *Computer Music Journal*, *18(1)*, 70-80.

Biographies

Godfried T. Toussaint has been doing research on ethnomusicology, music cognition, and computational music theory for the past ten years, with an emphasis on African rhythm. In 2005 he became a Researcher in the Music Department at McGill University (CIRMMT). From 2009 to 2010 he was the Emeline Bigelow Conland Fellow in the Radcliffe Institute for Advanced Study at Harvard University doing research on the phylogenetic analysis of the rhythms of the world. In September 2010 he joined the Music Department at Harvard University as a Research Scholar to carry out a research project on music cognition. It was at

Harvard University that this research was carried out.

Luke Matthews holds a B.S. in Biology and B.A. in Anthropology from Indiana University of Pennsylvania. He went on to graduate studies in physical anthropology at New York University, from which he recieved an M.A. and Ph.D. He was a postdoc in the Department of Human Evolutionary Biology at Harvard University at the time this research was conducted. He is now the Director of Data Analysis at Activate Networks, Inc. Dr. Matthews' research interests center on cultural diffusion and evolution in humans and nonhuman primates. His dissertation focused on social learning and foraging traditions in white-fronted capuchin monkeys living in Amazonian Ecuador. His postdoctoral and recent work have integrated quantitative social network and phylogenetic models to better understand cultural evolutionary dynamics in humans.

Malcolm Campbell graduated from Harvard University with a Bachelors degree in Chemistry and Physics, Summa cum laude in 2010, and earned a Masters degree in Jazz Piano Performance from the New England Conservatory in 2011. He now works for the Kohane lab at Children's Hospital Boston studying Autism genetics.

Naor Brown is an undergraduate student at Harvard University studying Applied Mathematics with Computer Science and Economics. An avid guitarist, Naor is interested in mathematically modeling the qualitative and social sciences, in his research.