

DANCE MUSIC CLASSIFICATION USING INNER METRIC ANALYSIS

A Computational Approach and Case Study Using 101 Latin American Dances and National Anthems

Elaine Chew, Anja Volk (Fleischer) and Chia-Ying Lee*

University of Southern California Viterbi School of Engineering

Integrated Media Systems Center

Epstein Department of Industrial and Systems Engineering

3715 McClintock Avenue GER240 MC:0193, Los Angeles CA90089-0193, USA.

{echew, avolk, leechiy}@usc.edu

Abstract This paper introduces a method for music genre classification using a computational model for Inner Metric Analysis. Prior classification methods focussing on temporal features utilize tempo (speed) and meter (periodicity) patterns and are unable to distinguish between pieces in the same tempo and meter. Inner Metric Analysis reveals not only the periodicity patterns in the music, but also the accent patterns peculiar to each musical genre. These accent patterns tend to correspond to perceptual groupings of the notes. We propose an algorithm that uses Inner Metric Analysis to map note onset information to an accent profile that can then be compared to template profiles generated from rhythm patterns typical of each genre. The music is classified as being from the genre whose accent profile is most highly correlated with the sample profile. The method has a computational complexity of $O(n^2)$, where n is the length of the query excerpt. We report and analyze the results of the algorithm when applied to Latin American dance music and national anthems that are in the same meter (4/4) and have similar tempo ranges. We evaluate the efficacy of the algorithm when using two variants on the model for Inner Metric Analysis: the metric weight model and the spectral weight model. We find that the correct genre is either the top rank choice or a close second rank choice in almost 80% of the test pieces.

Keywords: Music information processing, genre classification, rhythmic similarity.

*The research has been funded in part by the Integrated Media Systems Center, an NSF ERC, Cooperative Agreement No.EEC-9529152 and USC's WISE program. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of NSF or WISE.

Introduction

This paper proposes a method for genre classification by automatic extraction of rhythmic features from digital music. Accurate genre classification is critical to the efficient retrieval of music information. Previous research in genre classification using temporal information utilized tempo (speed) and meter (periodicity) as distinguishing features. A problem results when such classification methods encounter music of different genres that have the same tempo and meter, not an infrequent occurrence. We propose an algorithm for automatic classification that is capable of making subtle distinctions between different musical genres that can be in the same meter and tempo. The algorithm can be used either alone or as a add-on to existing classification methods. We test the algorithm using selected Latin American dances – the tango, the rumba, the bossa nova and the merengue – and national anthems. All test pieces have the same 4/4 meter, that is to say, they exhibit a periodicity pattern that cycles every four beats, and are typically played in a moderate tempo. Classification techniques that use tempo and periodicity information would not be able to distinguish between these pieces. The difference between these dances lie in the placement of strong and weak pulses within the four-beat framework. The entire test set consists of 101 pieces. We tested two versions of the algorithm that compute the accent profiles using two variations on the Inner Metric Analysis model [2, 5, 7], namely, the metric weight model and the spectral weight model. The correct genre was ranked either first or second in close to 80% of the test pieces. The metric weight model ranked the correct genre first in 57 of the test pieces, and within the top two rankings in 78 of the test pieces. The corresponding results for the spectral weight model were 71 and 79 respectively.

The perception of musical rhythm is one of the most basic human perceptual facilities. Music consists of a succession of sounds that can be represented as sequences of events that vary over time. The human ear possesses the innate ability to group these events into beats, units of time marking the rhythmic pulse of the music; the beats, in turn, coalesce into larger groups that determine the higher level metrical structures that are marked by patterns of strong (accented) and weak beats. Several other researchers have used temporal information to perform genre classification. This body of work use predominantly the beat and tempo (speed) information to differentiate between pieces from distinct genres. For example, Tzanetakis and Cook [12] use a combination of tempi, pitch histograms and sound texture to perform genre classification in audio files. They use beat histograms to show the predominant tempi in a musical audio excerpt. Tzanetakis and Cook reported a 61% accuracy rate in distinguishing between ten classes of music ranging from classical to metal, with classical and jazz having four and six subclasses respectively. Dixon,

Pampalk and Widmer [1] used a combination of tempo ranges and periodicity patterns (meter) to perform genre classification. They reported an 80% accuracy rate in distinguishing between fourteen styles of ballroom dance music. Gouyon and Herreraa [4] report a high degree of success (between 81 and 95%) in distinguishing between duple and triple meters. Other work on the use of periodicity patterns in similarity assessment includes Foote, Cooper and Nam's results on using self-similarity to extract periodicity patterns [3]. The use of periodicity patterns enabled finer grain distinction between dance music in different meters and improved upon tempo-only approaches to classification in the temporal domain. However, these methods that use only periodicity and tempo information encounter problems classifying music that are from different genres but exhibit the same periodicity and tempo characteristics. Much information is lost by not considering the structures encoded in the temporal patterns exhibited in the pieces themselves.

Our method focuses on extracting and comparing the grouping patterns induced by the note onsets present in the music. We employ the computational method of the Inner Metric Analysis described in [2], [5] and [7] to extract these temporal groupings from note onset time information and map them to time series of impulses that chart the relative strength of the pulse at each point in time. The method uses persistent and regular pulses at all grid levels and phase shifts to induce the metrical patterns present in the piece, and can be applied to both melodic and polyphonic music. We use tempo-invariant numeric input to bypass the problems introduced by performance tempo fluctuations and focus on the problem of perceptual grouping of pulses as exhibited in the piece. We compare the resulting accent profile for each piece of music to that of template rhythms typical of each musical genre by calculating the correlation coefficient between the two time series. Note that periodicity information will emerge in the resulting profiles. The music is classified as being from the genre whose template produces the highest correlation value. Distinct from some previous approaches, our classification method does not require pitch or tempo information.

The paper is organized as follows: Section 1 introduces the computational model for Inner Metric Analysis. Section 2 describes the algorithm for quantifying rhythmic similarity and presents the prototypical rhythm templates and details of the test set. Section 3 reports our computational results. Sections 3.2 and 4 presents a discussion of the results and conclusions respectively. A quick overview of the rhythm notation used in this paper can be found at www-rcf.usc.edu/~echew/papers/ICS2005.

1. Inner Metric Analysis

We use the models for Inner Metric Analysis described by Volk in [2] and [13] to map note onsets to numeric time series that reflect their accent strength. In traditional Western music notation, the grouping of beats is prescribed by a time signature in the musical score that indicates the period of the cycle and the unit of the beat. We call the time structure imposed by the time signature the *outer meter*. Each time signature imposes a particular accent pattern on the notes in each cycle that forms the rhythmic feature of the piece. Another type of time structure can be induced from the notes of the piece, by the groupings that arise out of the note content itself – we call this the *inner meter*. The two types of metrical structures may not always coincide. While the outer meter usually remains constant in a piece of music, the inner meter can shift dynamically according to the local grouping patterns. The correspondence, or lack thereof, between the *outer* and *inner meter* led to Volk’s definition of metric coherence [2, 13], which is said to exist if the the patterns in the inner meter exhibit the same period and phase as that of the outer meter. The investigation of a wide range of compositions of different styles and epochs concerning the occurrence of metric coherence has proven that it serves as an adequate description of the relation of the metric structure expressed by the notes of a piece and the outer meter. In this section, we give an overview of two methods for computing the inner meter from note onset information only. In Sections 1.1 and 1.2, we show how the inner metric structure can be generated by means of the *metric* and *spectral* weights respectively. More detailed descriptions of the basic concepts of Inner Metric Analysis and a software implementation can be found in [5], [2], [7] and [11].

1.1 Metric Weights

The objective of the model for metrical coherence is the mapping of a numeric weight to each note. The main concept behind the metric weight is the idea of a *local meter*, a maximal and successive chain of equally spaced events. If X denotes the set of all note onsets in a piece of music, then a local meter is defined as a maximal subset:

$$m(s, d, k) = \{s + id, i = 0, \dots, k\} \subset X,$$

where s denotes the starting point or first onset, d the fixed distance between the onsets of m or period, k the length of m . The local meter, $m(s, d, k)$, is maximal in that there does not exist another local meter $m' \subset X$ s.t. $m \subset m'$.

Figure 1 shows all local meters for a short excerpt from “The Girl From Ipanema” by Antonio Carlos Jobim, a well known example of the bossa nova genre. The discs on the first line below the notes (labeled X) show all onsets in the excerpt. There are three local meters present in this excerpt, namely,

$m(6, 2, 2)$, $m(6, 4, 2)$ and $m(6, 8, 2)$. These three local meters are indicated as successions of dark discs (●) in the second (A), third (B) and fourth (C) lines respectively. The triangles (△) show the extensions of the local meters, to be discussed in the next section.

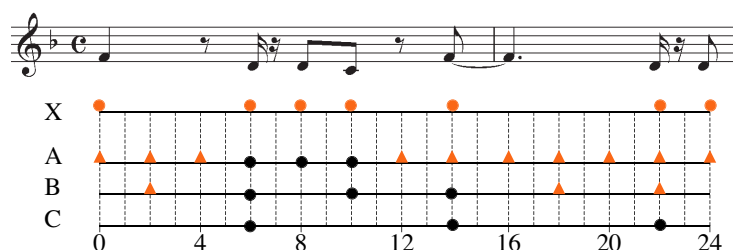


Figure 1. Local Meters in excerpt from “The Girl From Ipanema.”

The intuition behind the computation of the metric weight is that longer successions of events should contribute more weight to an onset in its chain than shorter ones. The more stably established in a chain of successive and regular events, the more significant the onset. Hence, the *metric weight* of an onset, $o \in X$, is the sum of a power function of the lengths of all local meters of size at least ℓ , of which o is an element. Let $M(\ell)$ be the set of all local meters of size at least ℓ , that is to say, $M(\ell) = \{m(s, d, k) : k \geq \ell\}$. The general metric weight of an onset, $o \in X$, is as follows:

$$W_{\ell,p}(o) = \sum_{\{m \in M(\ell) : o \in m\}} k^p. \quad (1)$$

For the purposes of this paper, we shall use the metric weight function when $\ell = 2$ and $p = 2$. For example, the second note in the excerpt in Figure 1 is a member of three local meters – $m(6, 2, 2)$, $m(6, 4, 2)$ and $m(6, 8, 2)$ – each of which has a length of two segments. Hence, the metric weight of this second note is $2^2 + 2^2 + 2^2 = 12$. Whereas, the third note is only a member of one local meter, $m(6, 2, 2)$, and its metric weight is $2^2 = 4$. Figure 2(a) shows the metric weight profile for all notes in the excerpt of “The Girl From Ipanema” shown in Figure 1. The higher the line in the graph, the greater the corresponding weight.

Note that if n is the length of the excerpt to be analyzed, all local meters at a given grid level d , $m(*, d, *)$, can be located in $O(n)$ time. There are n/ℓ such grid levels that are of interest. Hence, the computational complexity of the method for generating the metric weights is $O(n^2)$. The same is true for the spectral weight method described in the section to follow.

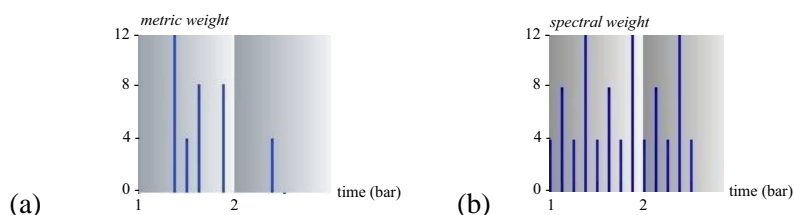


Figure 2. Metric and spectral weight profiles for “The Girl From Ipanema” excerpt.

1.2 Spectral Weights

The *spectral weight* method [7] extends the metric weight approach. In the metric weight approach, each local meter only contributes to the weight of onsets that belong to that local meter. In the spectral weight approach, each local meter contributes to the weight of onsets in its extension throughout the piece. The spectral weight exists not only for note events, but also for silence events (rests), on the other hand, it is not as sensitive to local changes in the inner meter as the metric weight.

Let the *extension* of a local meter be $ext(m(s, d, k)) = \{s + id, \forall i\}$. In Figure 1, the union of all symbols (circles, \bullet , and triangles, \triangle) on each line (A, B and C) shows the elements in the extensions of the corresponding local meters – $ext(m(6, 2, 2))$, $ext(m(6, 4, 2))$ and $ext(m(6, 8, 2))$. Each local meter, $m(s, d, k)$, contributes to the spectral weight of the events in its extension, $t \in ext(m(s, d, k))$. The spectral weight is defined as:

$$SW_{\ell,p}(t) = \sum_{\{m \in M(\ell): t \in ext(m)\}} k^p. \quad (2)$$

The metric and spectral weights for the “Girl From Ipanema” excerpt are given in Figures 2(a) and (b) respectively. While the metric weight profile does not reveal a regular grouping for this short example, the spectral weights can be separated into three hierarchical pulse levels, corresponding to the strongest pulses at the fourth and eighth elements in the bar, a moderate pulse on the second and sixth elements in the bar, and the underlying pulse on the beat.

2. Measuring Rhythmic Similarity

The comparison algorithm is based on the assumption that rhythms that are similar will generate similar distributions of strong and weak accent patterns on the time line as represented by their metric and spectral weight profiles. Section 2.1 outlines our method for genre classification. The test corpus is discussed in Section 2.2 and the prototypical rhythms for each class of music are presented in Section 2.3. An example of the classification process in action

is given in Section 2.4. The actual classification results for the test corpus will be presented in Section 3.

2.1 Quantifying Similarity

The core classification process involves two stages: the pre-processing of template rhythms representative of each class of dance music, and the classification of musical samples. In the first stage, we use template rhythms that are known to be typical of each genre to generate metric or spectral weight profiles that are representative of the class of music. The template rhythms are lifted from “The Complete Book of the World’s Dance Rhythms” by Kleon Raptakis [9] and augmented with examples gleaned from the CD “The Fabulous Ballroom Collection” [6].

In the calculation of the metric and spectral weights, local meters must be at least of length ℓ to contribute to the weight of an event. As a result, the first (and last) few events in a sample are assigned low weights that may not correspond to their true accent strengths. To minimize this edge effect and to get a stable accent profile, each rhythm pattern is repeated four times and a sample profile from the middle of this test set is singled out as the prototype. In general, the profiles for the second and third repetitions are highly similar. In this paper, we use the profile of the second repetition. The following steps are applied to each template rhythm:

Stage 1: Pre-processing of template rhythms

Step 1: Repeat the template rhythm four times.

Step 2: Calculate its metric/spectral weight.

Step 3: Extract metric/spectral weight profile of 2nd repetition.

Step 4: Normalize all weights to fall between 0 and MAX (=10).

The template rhythms and their accent profiles will be described in Section 2.3.

The second stage involves the comparison of each sample’s metric or spectral weight profile to the corresponding profiles of the prototypical rhythms. A short melodic fragment is excerpted manually near the beginning of the test sample and treated in the same fashion as the template rhythms (repeated four times, and middle of accent profile excerpted for comparison). We found that using short melodic fragments from a single instrument works well enough for classification purposes. The constant and regular succession of onsets typical of many Latin American pieces results in a regular onsets that do not reveal much accent differentiation when one considers the union of onsets from all instruments. The classification stage consists of the following steps:

Stage 2: Classification of music sample

Step 1: Obtain melodic fragment near beginning of sample.

Step 2: Repeat the rhythm pattern four times.

Step 3: Calculate its metric/spectral weight.

Step 4: Extract metric/spectral weight profile of 2nd repetition.

Step 5: Normalize all weights to fall between 0 and MAX (=10).

Step 6: Extend (by repetition) both the sample profile and the template profile to the lowest common multiple of sample and template lengths.

Step 7: Compute the correlation coefficient between the (extended) sample and template profiles (details to follow).

Step 8: Assign piece to genre with highest correlation.

Suppose that we wish to compare two accent profiles of the same length, $\{x_i : i = 1, \dots, N\}$ and $\{y_i : i = 1, \dots, N\}$, where N is large enough to represent all events in the accent profile. $x_i = 0$ if there is no accent weight at that point in time; the same is true for y_i . We use the correlation coefficient formula: $r = \sum x_i y_i / \sqrt{\sum x_i^2} \sqrt{\sum y_i^2}$.

2.2 Test Corpus

Our test corpus consists of twenty-two tangos, seventeen rumbas, twenty-five bossa novas, seventeen merengues and twenty national anthems from Latin America. All rhythms in the test corpus are of the same periodicity, 4/4 time, cycling after four beats; and, all except the merengue are typically played in a moderate tempo. Hence, the distinguishing features lie in the rhythm patterns themselves and the instrumentation, and not the periodicity or tempi.

The sources for our tango, rumba, bossa nova and merengue test data are listed online at www-rcf.usc.edu/~echew/papers/ICS2005/appendix.html. The source for the anthems was Reed and Bristow's "National Anthems of the World" [10]. We use as input to the classification process note material from a melodic fragment near the beginning of each piece. These melodic fragments were extracted manually and the duration sequences encoded in a text format. In the case of tangos, merengues, bossa novas and anthems, fragments of the melodies were excerpted as input. In the case of the rumba, the excerpt generally came from the bass instrument.

2.3 Prototypical Rhythms

We present here the prototypical rhythms for the tango, rumba, bossa nova and merengue. These rhythms were collected from the "Complete Book of the World's Dance Rhythms" [9] and augmented by templates manually transcribed from the CD "The Fabulous Ballroom Collection" [6].

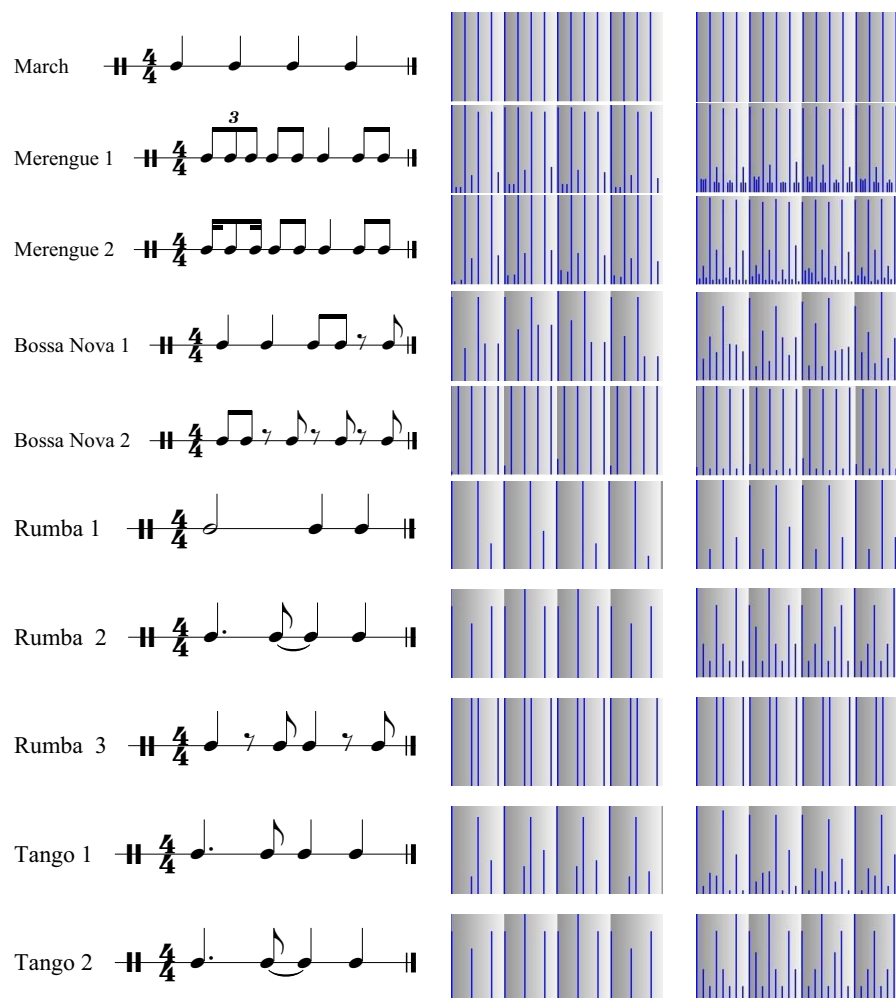


Figure 3. Prototypical rhythms with their respective metric (left chart) and spectral (chart on right) weight profiles.

Figure 3 documents the list of template rhythms for the musical genres under consideration. Each profile is derived from four repetitions of the rhythm shown. The first template is for the four-beat March rhythm typical of many national anthems. The typical merengue rhythms show a busy flow of notes. The bossa nova rhythms are characterized by eighth notes on the off-beats. The phase shift of the accent pattern to the off beats is particularly apparent in the profiles for the second bossa nova rhythm. Rumba 1 and Rumba 3 have strong accents on the first and third beats, as reflected in their metric and spectral weight profiles. Rumba 3 has an equally strong accent on the second (synco-

pated) note. The second rumba rhythm is identical to Tango 2. The difference between Rumba 2 and Tango 2 lies beyond the rhythmic realm, in the instrumentation and the pitch ornamentation employed in the pieces. For example, the use of the bandoneon in the tango but not the rumba, and the tango is punctuated by glissando-like figures that lead into important downbeats, the same is not true for the rumba. The first template rhythm for the tango is that of the European tango while the second is that of the Argentinian tango rhythm. The two rhythm patterns differ only in the third note, which is sounded in the European tango (creating a march-like rhythm) but is silent in the Argentinian version. The Argentinian tango rhythm has a strong syncopation, an accent on a weak beat that offsets the regularity of typical four-beat rhythm.

2.4 Classification Example

In our classification procedure, each sample piece is compared to all prototypical rhythms and assigned to the genre whose rhythm best matches its own as measured by the highest correlation coefficient value. We return at this point to “The Girl From Ipanema” (from Figure 1) to demonstrate the classification procedure. The excerpted rhythm corresponds to the lyrics “Tall and tan and young and love-”. This excerpt is repeated four times and its metric and spectral weight profiles computed and shown in Figure 4. The profile of the second

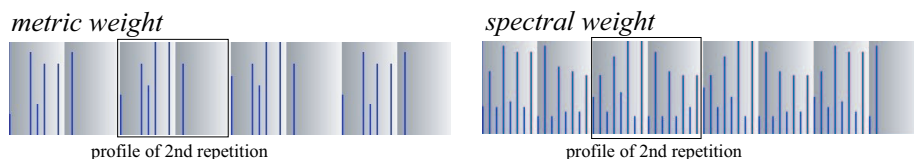


Figure 4. Accent profiles for four repetitions of “The Girl From Ipanema” excerpt.

repetition is excerpted for comparison to all rhythm templates (see Section 2.3). We shall refer to this as the Ipanema profile. Each rhythm template (the profile of the second repetition of each template rhythm in Figure 3) is repeated twice so that it is of the same length as the Ipanema example. We then compute the correlation coefficient between the Ipanema profile and the extended rhythm templates. The coefficient correlation values for all rhythm templates using both metric and spectral weights are presented in Table 1 (where A = anthem, B = bossa nova, M = merengue, R = rumba and T=tango).

Both the metric weight and the spectral weight approaches rank the Bossa nova 2 rhythm highest, followed by the Rumba 3 rhythm. Consider the Bossa nova 2 rhythm in Figure 3. Both the Bossa nova 2 template and the Ipanema

Table 1. Correlation values for comparison of Ipanema profile to rhythm templates.

	B2	R3	B1	T1	R1	M2	M1	T2	R2	A
metric wt	0.62	0.55	0.49	0.41	0.36	0.34	0.34	0.31	0.31	0.26
spectral wt	0.62	0.61	0.44	0.33	0.37	0.45	0.54	0.31	0.31	0.30

profiles (metric and spectral weights) display prominent weights on the off beats, thus identifying the Ipanema rhythm as being from the bossa nova genre.

3. Computational Results

This section reports our computational results using the method outlined in Section 2.1, the test corpus detailed in Section 2.2 and the rhythm templates described in Section 2.3. Section 3.1 describes the experiments and the classification results and Section 3.2 provides a discussion of the results.

3.1 Classification Experiments and Results

We performed separate classification tests using the metric weight model and the spectral weight model. The summary statistics on the correlation coefficient values are reported in Table 2.

Table 2. Summary statistics for correlation coefficient values.

<i>Rhythm profile method</i>	<i>Highest corr coeff average (std dev)</i>	<i>Lowest corr coeff average (std dev)</i>
metric weight	0.7676 (0.1784)	0.2058 (0.1338)
spectral weight	0.7209 (0.1804)	0.2456 (0.1268)

We first examine the rank one assignment results, that is to say, those with the highest correlation coefficient scores. The left and right charts and tables in Figure 5 document the rank one classification results using the metric and spectral weight models respectively. The bar chart shows the classification results for each test sample class. Each table is a transformed confusion matrix that maps directly to the segments in its corresponding bar chart. The numbers in bold indicate the correct assignments. In infrequent cases of ambiguity, a piece is considered to have been correctly classified if the rhythm template with which its melodic fragment achieved the highest correlation coefficient value is one of the prototypical rhythms of its genre. Consider the identical rhythm templates: Rumba 2 and Tango 2. If a rumba matches this rhythm template, then it is considered to have been classified correctly.

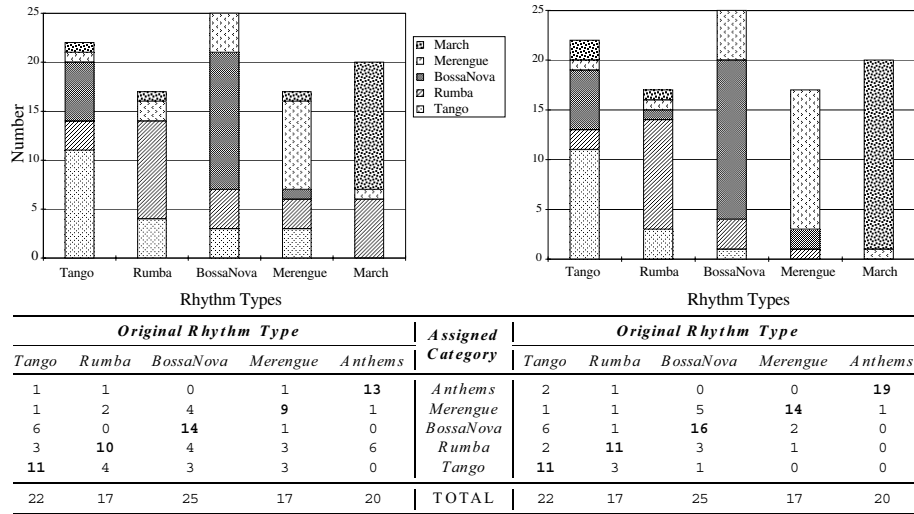


Figure 5. Rank one categorization results using metric (left) and spectral (right) weight models.

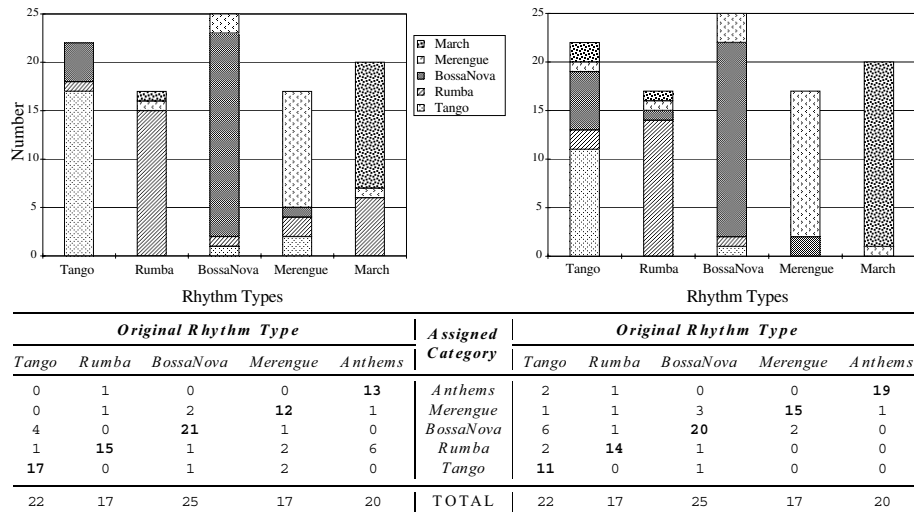


Figure 6. Rank one and two (within 0.05) categorization results using metric (left) and spectral (right) weight models.

Sometimes, a piece may be misclassified based on the top ranked assignment, but a close second rank competitor may be the correct genre assignment. Hence, we next take into account second rank results that are correct and close to the top rank classifications. If the correlation coefficient corresponding to the rank two classification of the sample is within, say, 0.05 of that of the rank one classification, then the second classification is considered one of the top choices. For this second scenario, the figures and tables in Figure 6 document the corresponding classification results using the metric and spectral weight models. Finally, Table 3 summarizes the classification results for each method.

Table 3. Summary of classification results.

<i>Experiment</i>	<i>Proportion correct: only rank one assignments</i>	<i>Proportion correct: rank one & two assignments (rank two within 0.05 of one)</i>
metric weight	0.5644 (57 out of 101)	0.7723 (78 out of 101)
spectral weight	0.7030 (71 out of 101)	0.7822 (79 out of 101)

3.2 Discussion and Analysis

In this section, we evaluate the results of our approach for fine-grain differentiation of the musical genres employing the Inner Metric Analysis model presented in Section 1.

Performance Summary. In general, the method using the spectral weights accurately classified more of the test pieces than the one using metric weights. Consider the top-ranked assignment results presented in Figure 5. The spectral weight approach provided more accurate classifications – 70.3% correct versus the metric weight’s 56.4% – showing a marked improvement in the accurate classification of the merengue and march (anthem) samples. For example, the six anthems that were incorrectly classified as rumbas (Rumba 1) using the metric weight approach (see chart on left in Figure 5) were adjusted to the right category in the spectral weight approach (chart on right in same figure). The metric weight approach performed markedly better than the spectral weight method only in assigning the correct classification of tango within the first and second choices (see the two tables at the bottom of Figure 6).

In a significant portion of cases for the metric weight approach where the first rank answer was incorrect, a close second rank answer was the correct one. The same phenomenon was observed to a lesser degree for the spectral weights approach. Compare the metric weight experiment results documented in the left charts in Figures 5 and 6. In 21 out of the 44 incorrect genre assignments

for the dance pieces, the correct assignment was the rank two answer whose correlation coefficient was within 0.05 of the rank one answer. For example, all four rumba pieces misclassified as tangos (Tango 1) had as second highest scoring template a rumba rhythm (Rumba 1 and Rumba 2). In fact, when considering both the highest and close second highest correlation coefficient scores, the performance of both the metric weight and spectral weight methods were comparable as shown in Table 3.

Table 4a. Most frequent classification errors using metric weight method.

<i>Original type</i>	<i>Probable orig.type (#)</i>	<i>Misclass. type (#)</i>
Anthem	Ma(6)	R1(6)
Tango	T1(4)	B1(4)
Rumba	R1(3), R2(1)	T1(4)
BossaNova	B1(1), B2(3)	R3(4)
BossaNova	B1(3)	M2(3)
Tango	T1(1), T2(1)	B2(2)
Merengue	M2(2)	T2(2)
Anthem	Ma(1)	M2(1)
BossaNova	B1(1)	M1(1)
Merengue	M1(1)	T1(1)

Table 4b. Most frequent classification errors using spectral weight method.

<i>Original orig.type</i>	<i>Probable type (#)</i>	<i>Misclass. type (#)</i>
Tango	T1(2), T2(2)	B1(4)
BossaNova	B1(4)	M1(4)
Rumba	R2(3)	T1(3)
Tango	T1(2)	B2(2)
Merengue	M1(2)	B2(2)
BossaNova	B1(1)	M2(1)
Anthem	Ma(1)	R1(1)

LEGEND: B = bossa nova, Ma = march
M = merengue, R = rumba, T = tango

Most Frequent Errors. We next consider the most frequent classification errors. Referring to Figure 5, the largest segment in each bar in the graph corresponds to the correct assignment, while the second largest segment represents the most frequent classification error. The details of these most frequent classification errors for the metric weight and spectral weight methods are summarized in Tables 4a and 4b respectively. Each table lists the most frequent erroneous assignment for each genre as well as their probable correct type. The probably type is the template in the correct genre that has the highest correlation with the sample's accent profile. For example, of the four rumbas misclassified as Tango1 in the metric weight table, three matched the Rumba1 (R1) rhythm and one the Rumba2 (R2) template. In order to assess the degree of error for misclassified samples, we present the similarity matrix for the template rhythms in Table 5. The tables show the similarity matrix for the template rhythms computed using the metric and spectral weights respectively.

The most frequent error using the metric weight profiles was the misclassification of anthems (marches) as rumbas (Rumba 1, R1, rhythm). This is not a surprising outcome since one might expect the R1 rhythm template to be a plausible march rhythms. In his experiments on pulse salience and metrical ac-

Table 5. Similarity matrices for rhythm templates.

<i>metric weight method</i>											<i>spectral weight method</i>										
	T1	T2	R1	R2	R3	B1	B2	M1	M2	Ma	T1	T2	R1	R2	R3	B1	B2	M1	M2	Ma	
T1	x	.67	<u>.97</u>	.67	.78	.76	.18	.82	.79	.78	x	<u>.99</u>	.62	<u>.99</u>	.51	.19	.42	.70	.63	.58	
T2		x	.48	<u>1.0</u>	.60	.32	.37	.61	.61	.53		x	.55	<u>1.0</u>	.49	.17	.45	.69	.61	.51	
R1			x	.48	.69	.79	.06	.80	.77	.80			x	.55	.58	.49	.23	.47	.55	.86	
R2				x	.60	.32	.37	.61	.61	.53				x	.49	.17	.45	.69	.61	.51	
R3					x	.71	.54	.64	.64	.50					x	.62	.66	.88	.74	.46	
B1						x	.33	.82	.81	.81						x	.52	.64	.64	.70	
B2							x	.18	.20	.04							x	.66	.51	.18	
M1								x	<u>.99</u>	<u>.98</u>								x	.84	.55	
M2									x	<u>.97</u>									x	.64	
Ma										x										x	

cent, Parncutt used the R1 rhythm to represent a march [8]. The second most frequent error was the misclassification of tangos as bossa novas (B1), rumbas as tangos (T1), and bossa novas as rumbas (R3). Consider the accent profile of the T1 and the B1 rhythms in Figure 3: they both show strong accents on the first and third beats, a high degree of correspondence with the outer meter, which may explain the algorithm’s confusion between the T1 and B1 classification. The correlation between the R1 and T1 rhythms is high (0.97 as shown in the left similarity matrix in Table 5); and, both the B2 and R3 rhythms show strong accents on the weak beats (see Figure 3).

The most frequent error using the spectral weight profiles was the misclassification of tangos as bossa novas (B1) and bossa novas as merengues (M1). As mentioned in the above paragraph, both the tango and B1 rhythms show strong accents on the first and third beats. Both the B1 and M1 rhythms exhibit strong accents on beats 1, 2 and 3.

4. Conclusions

We have presented an $O(n^2)$ method for music genre classification using a computational model for Inner Metric Analysis. We have shown that the inner metrical patterns of the music provide valuable information that can make fine distinctions between pieces in the same tempo and meter. We tested the algorithm on 101 Latin American dances and national anthems and ranked the correct genre within the top two choices in close to 80% of the test pieces.

The current implementation of the algorithm utilized manual preprocessing of the data to extract a melodic fragment for analysis. Future implementations could automate this process, introduce random selection of input data or classification by majority vote from all tracks, and explore the sensitivity of the algorithm to the choice of input fragments. While we have used score-generated MIDI input to encode musical information in order to validate the approach using data that is as precise as possible, further work needs to be

done to modify the algorithm to perform classification under uncertainty, for example, irregularity of beats as would be the case in performed music.

In conclusion, the classification method we have presented can discriminate between musical genres based on accent patterns in the same meter (periodicity). Such differentiation by accent patterns provided by Inner Metric Analysis can be used either as a standalone technique or in conjunction with tempo information as a method for genre classification that employs a more complete set of temporal features.

References

- [1] Dixon, Simon, Pampalk, Elias and Widmer, Gerhard (2003). "Classification of Dance Music by Periodicity Patterns," in Proceedings of the 4th International Conference on Music Information Retrieval, Baltimore, MD, p. 159-165.
- [2] Fleischer, Anja (2002). "A Model of Metric Coherence," in Proceedings of the 2nd International Conference on Understanding and Creating Music, Caserta, Italy.
- [3] Foote, Jonathan, Cooper, Matthew and Nam, Unjung (2002). "Audio Retrieval by Rhythmic Similarity," in Proceedings of the 2002 International Conference on Music Information Retrieval at IRCAM, Paris.
- [4] Gouyon, Fabien, and Herrera, Perfecto (2003). "Determination of the meter of musical audio signals: Seeking recurrences in beat segment descriptors" in Proceedings of the 114th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, 2003.
- [5] Mazzola, Guerino (2002). *The Topos of Music*. Basel: Birkhauser.
- [6] Murray, Arthur (1998). *The Fabulous Ballroom Collection*. RCA CD No.63136 (UPC: 90266313624)
- [7] Nestke, Andreas and Noll, Thomas (2001). "Inner Metric Analysis'," in Haluska, Jan (ed.) *Music and Mathematics*. Bratislava: Tatra Mountains Mathematical Publications.
- [8] Parncutt, Richard (1994). "A Perceptual Model of Pulse Saliency and Metrical Accent in Musical Rhythms," *Music Perception*, 1994, Vol. 11, No. 4, p. 409-464.
- [9] Raptakis, Kleon (1966). *The Complete Book of the World's Dance Rhythms*. Astoria, N.Y.: K. Raptakis Publications.
- [10] Reed, William L. and Bristow, Michael J. (1987). *National Anthems of the World*. 7th ed. ISBN: 0713719621. London; New York: Blandford Press.
- [11] Volk, Anja and Noll, Thomas (2004). *Rubato, Rubettes and related Java-Tools: A User's Guide*. Manuscript.
- [12] Tzanetakis, George and Cook, Perry (2002). "Musical Genre Classification of Audio Signals," *IEEE Transactions on Speech and Audio Processing*, 10(5), July 2002.
- [13] Volk, Anja (2004). *Metric Investigations in Brahms' Symphonies*. In Mazzola, Guerino and Thomas Noll (ed): *Perspectives of Mathematical and Computer-Aided Music Theory*, epOs Music, Osnabrück.