BEAT AND DOWNBEAT TRACKING BASED ON RHYTHMIC PATTERNS APPLIED TO THE URUGUAYAN CANDOMBE DRUMMING

Leonardo Nunes ATL-Brazil, Microsoft lnunes@microsoft.com

Martín Rocamora **Luis Jure** Universidad de la República {rocamora,lj}@eumus.edu.uy

Luiz W. P. Biscainho Federal Univ. of Rio de Janeiro wagner@smt.ufrj.br

ABSTRACT

Computational analysis of the rhythmic/metrical structure of music from recorded audio is a hot research topic in music information retrieval. Recent research has explored the explicit modeling of characteristic rhythmic patterns as a way to improve upon existing beat-tracking algorithms, which typically fail on dealing with syncopated or polyrhythmic music. This work takes the Uruguayan Candombe drumming (an afro-rooted rhythm from Latin America) as a case study. After analyzing the aspects that make this music genre troublesome for usual algorithmic approaches and describing its basic rhythmic patterns, the paper proposes a supervised scheme for rhythmic pattern tracking that aims at finding the metric structure from a Candombe recording, including beat and downbeat phases. Then it evaluates and compares the performance of the method with those of general-purpose beat-tracking algorithms through a set of experiments involving a database of annotated recordings totaling over two hours of audio. The results of this work reinforce the advantages of tracking rhythmic patterns (possibly learned from annotated music) when it comes to automatically following complex rhythms. A software implementation of the proposal as well as the annotated database utilized are available to the research community with the publication of this paper.

1. INTRODUCTION

Meter plays an essential role in our perceptual organization of music. In modern music theory, metrical structure is described as a regular pattern of points in time (*beats*), hierarchically organized in metrical levels of alternating strong and weak beats [15, 16]. The metrical structure itself is not present in the audio signal, but is rather inferred by the listener through a complex cognitive process. Therefore, a computational system for metrical analysis from audio signals must, explicit or implicitly, make important cognitive assumptions. A current cognitive model proposes that, given a temporal distribution of events, a competent listener infers the appropriate metrical structure by applying two sets of rules: Metrical Well-Formedness Rules (MWFR), which define the set of possible metrical structures, and Metrical Preference Rules (MPR), which model the criteria by which the listener chooses the most stable metrical structure for a given temporal distribution of events [15]. While not strictly universal, most of the MWFR apply for a variety of metric musics of different cultures [23]; MPR, on the other hand, are more subjective and, above all, style-specific. A listener not familiar with a certain type of music may not be able to decode it properly, if its conventions differ substantially from usual tonal metrical structures.

This is why the computational analysis of rhythmic/metrical structure of music from audio signals remains a difficult task. Most generic algorithms follow a bottom-up approach with little prior knowledge of the music under analysis [6, 7, 13], often including some kind of preference rules—e.g. by aligning beats with onsets of stronger and/or longer events [15]. Therefore, they usually fail on processing syncopated or polyrhythmic music, for instance, that of certain Turkish, Indian or African traditions [22].

For this reason, other approaches prefer a top-down process guided by high-level information, such as style-specific characteristics [11]. Given that listeners tend to group musical events into recurrent rhythmic patterns which give cues for temporal synchronization, the explicit modeling of rhythmic patterns has recently been proposed as a way to improve upon existing beat-tracking algorithms [14, 24, 25]. The identification of challenging music styles and the development of sytle-specific algorithms for meter analysis and beat-tracking is a promising direction of research to overcome the limitations of existing techniques.

In this work, an afro-rooted rhythm is considered as a case of study: the Candombe drumming in Uruguay. Motivated by the fact that some characteristics of Candombe are challenging for most of the existing rhythm analysis algorithms, a supervised scheme for rhythmic pattern tracking is proposed, aiming at finding the metric structure from an audio signal, including the phase of beats and downbeats. The performance of the proposed method is assessed over a database of recordings annotated by an expert.

The next section provides a brief description of the Candombe rhythm. Then, the proposed method for rhythmic pattern matching is presented in Section 3. Experiments and results are described in Section 4. The paper ends with some critical discussion and directions for future research.

2. AFRO-URUGUAYAN CANDOMBE

2.1 Candombe drumming in context

Candombe is one of the most characteristic features of Uruguayan popular culture, practiced by thousands of people. Its rhythm influenced and was incorporated into various genres of popular music. However, little known abroad, it may be difficult to understand for unfamiliar listeners.

Although originated in Uruguay, Candombe has its roots in the culture brought by the African slaves in the 18th century. It evolved during a long historical process, gradually integrating European immigrants and now permeating the whole society [1, 8]. Candombe drumming, with its distinctive rhythm, is the essential component of this tradition. Its most characteristic manifestation is the *llamada de tambores*, a drum-call parade, when groups of players meet at specific points in the city to play while marching on the street (Figure 1).

Figure 1. Group of Candombe drummers.

The instrument of Candombe is called *tambor* (“drum” in Spanish), of which there are three different sizes: *chico* (small), *repique* (medium) and *piano* (big). Each type has a distinctive sound (from high to low frequency range) and its own specific rhythmic pattern. All three are played with one hand hitting the skin bare and the other with a stick, which is also used to hit the shell when playing the *clave* pattern. The minimal ensemble of drums (*cuerda de tambores*) must have at least one of each of the three drums; during a *llamada de tambores* it usually consists of around 20 to 60 drums. While marching, the players walk forward with short steps synchronized with the beat or *tactus*; this movement, while not audible, is very important for the embodiment of the rhythm. Figure 1 shows in the first row, from the front backwards, a *repique*, a *chico* and a *piano*.

2.2 Rhythmic patterns and metrical structure

The Candombe rhythm or *ritmo de llamada* results from the interaction between the patterns of the three drums. An additional important pattern is the *clave*, played by all the drums as an introduction to and preparation for the rhythm (see Figure 2 [1]).

The pattern of the *chico* drum is virtually immutable, and establishes the lowest level of the metrical structure (*tatum*). The period of the pattern is four *tatums*, conforming the beat or *tactus* level in the range of about 110 to 150 beats per minute (BPM). The interaction of *chico* and *clave* helps to establish the location of the beat within the *chico* pattern (otherwise very difficult to perceive), and defines a higher metric level of four beats (sixteen *tatums*).

Figure 2. Interaction of main *Candombe* patterns, and the three levels of the resulting metric structure. *Repique* and *piano* patterns are shown in a simplified basic form.

The resulting metrical structure is a very common one: a four-beat measure with a regular subdivision in 16 *tatums*. However, two characteristic traits link the rhythmic configuration of Candombe with the Afro-Atlantic music traditions, differentiating it from usual tonal rhythms: 1) the pattern defining the pulse does not articulate the *tatum* that falls on the beat, and has instead a strong accent on the second; 2) the *clave* divides the 16-*tatum* cycle irregularly (3+3+4+2+4), with only two of its five strokes coinciding with the beat. This makes the Candombe rhythm difficult to decode for both listeners not familiar with it and generic beat-tracking algorithms (see Table 1). The strong phenomenological accents displaced with respect to the metric structure add to the difficulty.

The *repique* is the drum with the greatest degree of freedom. During the *llamada* it alternates between the *clave* pattern and characteristically syncopated phrases. Figure 2 shows its primary pattern, usually varied and improvised upon to generate phrases of high rhythmic complexity [12]. The *piano* drum has two functions: playing the base rhythm (*piano base*), and occasional more complex figurations akin to the *repique* phrases (*piano repicado*). The pattern in Figure 2 is a highly abstracted simplification of the *piano base*. It can be seen that it is essentially congruent with the *clave* pattern, and when correctly decoded it permits the inference of the whole metric structure. In real performances, however, much more complex and varied versions of this pattern are played. It has been shown [21] that the analysis of *piano* patterns may elicit the identity of different neighborhoods (*barrios*) [2] and individual players.

3. RHYTHMIC PATTERN MATCHING

In this section, a rhythmic/metric analysis algorithm that matches a given rhythmic accentuation pattern to an audio signal is described. It tries to find the time of occurrence

[1] Lower and upper line represent hand and stick strokes respectively.

[2] The three more important traditional styles are *Cuareim* (or *barrio Sur*), *Ansina* (or *barrio Palermo*) and *Gaboto* (or *barrio Cordón*).

of each *tatum* knowing its expected accentuation inside the pattern, thus being able to track not only the beat but also other metrical information. Initially, a tempo estimation algorithm is employed to obtain the beat period (tempo), assumed to be approximately stable throughout the signal. Then, the main algorithm is used to find the phase of the accentuation pattern within the observed signal.

3.1 Audio feature extraction

For audio feature extraction, this work adopts a typical approach based on the Spectral Flux. First, the Short-Time Fourier Transform of the signal is computed and mapped to the MEL scale for sequential windows of 20 ms duration in hops of 10 ms. The resulting sequences are differentiated (via first-order difference) and half-wave rectified.

For tempo estimation, the feature values are summed along all MEL sub-bands, in order to take into account events from any frequency range.

Since its pattern is the most informative on both *tactus* beat and downbeat locations, the rhythmic pattern tracking is tailored towards the *piano* (i.e. the lowest) drum. Therefore, the accentuation feature used for pattern matching is obtained by summing the Spectral Flux along the lowest MEL sub-bands (up to around 200 Hz) only. This function is normalized by the 8-norm of a vector containing its values along ± 2 estimated *tatum* periods around the current frame. The resulting feature value is expected to be close to one if a pulse has been articulated and close to zero otherwise. In addition, it also carries some information on the type of articulation. For instance, an accented stroke produces a higher feature value compared to a muffled one, since in the former case the spectral change is more abrupt.

3.2 Tempo Estimation

For tempo estimation, this work adopts a straightforward procedure based on locating the maximum of a suitably defined similarity function. As proposed in [20], the basic function is the product between the auto-correlation function and the Discrete Fourier Transform of the features computed for the whole signal. The result is weighted by the function described in [17]. The period associated with the largest value in this weighted similarity function is selected as the tempo of the signal. After the tempo is obtained, the *tatum* period used for pattern tracking can be computed just by dividing the beat period by 4. This *tatum* period is then used to define the variables in the pattern tracking algorithm as described in the next sections.

3.3 Variables definition

In order to perform its task, the algorithm employs two discrete random variables. The first one, called *tatum counter*, $\mathbf{c}_k$, counts how many frames have passed since the last *tatum* has been observed at frame k. Assuming an estimated *tatum* period of τ frames, then $\mathbf{c}_k \in \{0, 1, \ldots, \tau - 1 + \sigma_c\}$, where σ_c is a parameter that allows for possible timing inaccuracies in the *tatum*. The second, called *pattern index*, $\mathbf{a}_k$, indicates the position inside a given rhythmic pattern at frame k in the range $\{0, 1, \ldots, M - 1\}$, where M is the length of the rhythmic pattern in *tatums*. The rhythmic pattern will be expected to define a series of accents or lacks of accent in the *tatums*. Time evolution of these two variables will be described in the next section, where it is assumed that the sampling rate of the feature (typically less than 100 Hz) is much lower than that of the original signal (usually 44.1 kHz). The model describes the accentuation feature extracted at frame k as a random variable, $\mathbf{y}_k$, with actual observed (extracted) value y_k.

3.4 State Transition

In this section, the probabilities of each value for the two random variables at frame k given past frames are described. A first-order Markov model will be assumed for the joint distribution of the random variables, i.e., the probability of each possible value of a random variable at frame k depends only on the values assumed by the variables at the previous frame $k - 1$. Using this assumption, the two random variables will constitute a Hidden Markov Model [18].

The *tatum* counter variable, as previously mentioned, counts how many frames have passed since the last *tatum*. The state $\mathbf{c}_k = 0$ is considered the "tatum state" and indicates that a *tatum* has occurred at frame k. This random variable is closely related to the *phase state* proposed in [5] for beat tracking. Only two possible transitions from frame $k - 1$ to frame k are allowed: a transition to the "tatum state" or an increment in the variable. The transition to the "tatum state" depends on both the past value of the variable and the (known) *tatum* period. The closer the value of the variable is to the *tatum* period, the more probable is the transition to the "tatum state." Mathematically, it is possible to write

$$p_{\mathbf{c}_k}(c_k|c_{k-1}) = \begin{cases} h[c_{k-1} - \tau], & \text{if } c_k = 0 \\ 1 - h[c_{k-1} - \tau], & \text{if } c_k = c_{k-1} + 1 \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

where $h[.]$ is a tapering window with $h[n] = 0$ for $|n| > \sigma_c$ that models possible timing inaccuracies on the *tatum*, and $\sum_n h[n] = 1$. Currently, a normalized Hann window is employed to penalize farther values. The value $\sigma_c = 2$ was set for the reported experiments, indicating that inaccuracies of up to 50 ms are tolerated by the algorithm.

Since the accentuation pattern is defined in terms of the *tatum*, its time evolution will be conditioned by the pattern evolution. Assuming that the pattern indicates the expected accentuation of the next *tatum*, the variable should only change value when a "tatum state" has been observed, indicating that a different accentuation should be employed by the observation model (described in the next section). Hence, mathematically

$$p_{\mathbf{a}_k}(a_k|c_{k-1}, a_{k-1}) = \begin{cases} 1, \text{ if } (a_k = a_{k-1} \oplus 1) \wedge (c_{k-1} = 0) \\ 1, \text{ if } (a_k = a_{k-1}) \wedge (c_{k-1} \neq 0) \\ 0, \text{ otherwise,} \end{cases} \tag{2}$$

where $\wedge$ is the logical AND, $\oplus$ denotes a modulo-M summation, and M is the length of the accentuation pattern. As can be gathered, given the previous *tatum* counter value, the pattern index becomes deterministic, with its next value completely determined by its value at the previous frame and the value of the *tatum* counter. The transitions for this variable are inspired on the ones used in the family of algorithms based on [24] (i.e. [2, 10, 14]), except for defining the pattern in terms of *tatums* instead of an arbitrary unit.

3.5 Observation Model

This section describes the likelihood of $\mathbf{c}_k$ and $\mathbf{a}_k$ given an observed accentuation $\mathbf{y}_k$ in the signal. The main idea is to measure the difference between the expected accentuation (provided by the rhythmic pattern) and the observed one. The larger the difference, the less probable the observation.

If the accentuation pattern is a vector $\overline{A} \in \mathrm{R}^{M\times 1}$ containing the expected feature values, then at frame k the likelihood for $c_k = 0$ ("tatum state") can be defined as

$$p_{\mathbf{y}_k}(y_k|c_k, a_k) = N_{\sigma_t}(y_k - \overline{A}_{a_k}), \tag{3}$$

where $N_{\sigma_t}(.)$ is a Gaussian function with zero mean and variance σ_t^2 used to model possible deviations between expected and observed accents. For $c_k \neq 0$, the likelihood is given by:

$$p_{\mathbf{y}_k}(y_k|c_k, a_k) = N_{\sigma_d}(y_k), \tag{4}$$

where N_{σ_d} is a zero-mean Gaussian with variance equal to σ_d^2. Hence, the closer to zero the feature, the more probable the observation. This is similar to the non-beat model adopted in [5], and is not found in [14, 24].

In the reported experiments, $\sigma_t = \sigma_d = 0.5$, thus allowing for a reasonable overlap between expected and actual observed values.

3.6 Inference

A summary of the proposed model for rhythmic pattern tracking can be viewed in Figure 3, where the statistical dependencies among the variables are explicited. Different inference strategies can be employed to find the most probable pattern index and *tatum* counter values given the observed accentuation [18]. In this work, the well-known Viterbi algorithm [18, 24] is employed to find the most probable path among all possible combinations of values of each random variable given the observed features y_k.

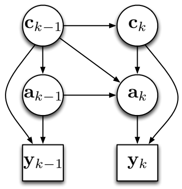

Figure 3. Graphical representation of the statistical dependency between random variables and observations.

At last, a uniform prior is chosen for $\mathbf{c}_0$ and $\mathbf{a}_0$ indicating that the counter and the pattern can start with any possible value in the first frame.

4. EXPERIMENTS AND RESULTS

A set of experiments was devised to assess the performance of the proposed rhythmic pattern tracking system with respect to the problems of estimating the rate and phase of beats and downbeats, using a database of manually labeled Candombe recordings. Four state-of-the-art beat-tracking algorithms [6, 7, 13, 19] were included in the experiments in order to evaluate how challenging the rhythm at hand is for typical general-purpose approaches.

Two different strategies are explored: the rhythmic patterns to follow are either informed to the algorithm based on a priori musical knowledge about the rhythm, or learned from the labeled database itself.

4.1 Dataset

A dataset of Candombe recordings, totaling over 2 hours of audio, was compiled and annotated for this work and it is now released to the research community. [3] It comprises 35 complete performances by renowned players, in groups of three to five drums. Recording sessions were conducted in studio, in the context of musicological research over the past two decades. A total of 26 *tambor* players took part, belonging to different generations and representing all the important traditional Candombe styles. The audio files are stereo with a sampling rate of 44.1 kHz and 16-bit precision. The location of beats and downbeats was annotated by an expert, adding to more than 4700 downbeats.

4.2 Performance measures

Since tempo estimation is only an initialization step of the rhythmic pattern tracking task, whose overall performance will be examined in detail, it suffices to mention that the estimated tempo was within the interval spanned by the annotated beat periods along each of the files in the database, thus providing a suitable value for the respective variable.

Among the several objective evaluation measures available for audio beat tracking [4] there is currently no consensus over which to use, and multiple accuracies are usually reported [2, 3]. In a recent pilot study, the highest correlation between human judgements of beat tracking performance and objective accuracy scores was attained for CMLt and Information Gain [3].

In this work CMLt, AMLt and F-measure were adopted, as their properties are well understood and were considered the most suitable for the current experiments. The non-inclusion of Information Gain was based on the observation that it yielded high score values for estimated beat sequences that were definitely not valid. Specifically, in several instances when the beat rate (or a multiple of it) was precisely estimated, even if the beat phase was repeatedly misidentified, the Information Gain attained high values while other measures such as CMLt or F-measure were coherently small. In the following, a brief description

[3] Available from `http://www.eumus.edu.uy/candombe/datasets/ISMIR2015/`.

of the adopted metrics [4] is provided (see [4] for details), along with the values selected for their parameters.

The CMLt measure (Correct Metrical Level, continuity not required) considers a beat correctly estimated if its time-difference to the annotated beat is below a small threshold, and if the same holds for the previous estimated beat. Besides, the inter-beat-interval is required to be close enough to the inter-annotation-interval using another threshold. The total number of correctly detected beats is then divided by the number of annotated beats and expressed as a percentage (0-100 %). Both thresholds are usually set to 17.5 % of the inter-annotated-interval, which was also the value adopted in this work. The AMLt measure (Allowed Metrical Levels, continuity not required) is the same as CMLt but does not take into account errors in the metrical level and phase errors of half the period.

The F-measure (Fmea) is the harmonic mean of precision and recall of correctly detected beats, where precision stands for the ratio between correctly detected beats and the total number of estimated beats, while recall denotes the ratio between correctly detected beats and the total number of annotated beats. A beat is considered correctly detected if its time-difference to the annotation is within ± 70 ms; this tolerance was kept in this work.

Only CMLt and F-measure were used for assessing the downbeat, since the loosening of metrical level and phase constraints in AMLt was considered inappropriate.

4.3 Experiments with informed rhythmic patterns

In the first type of experiment, the pattern to track $\overline{A}$ is informed to the algorithm based on musical knowledge about the rhythm, without any training or tuning to data. On one hand, this has a practical motivation: even when no labeled data is available one could take advantage of the technique. On the other hand, it gives a framework in which musical models can be empirically tested. In short, an informed rhythmic pattern based on musical knowledge is nothing but a theoretical abstraction, and this type of experiment could provide some evidence of its validity.

To that end, based on the different ways the *piano* pattern is notated by musicology experts, a straightforward approach was adopted. Firstly, the *piano* pattern as introduced in Figure 2 (usually regarded as the *piano* in its minimal form) was considered. A binary pattern $\overline{A}$ was assembled by setting a value of 1 for those *tatums* which are expected to be articulated and 0 otherwise. Then, a more complex pattern was considered by adding two of the most relevant articulated *tatums* which were missing, namely the 6th and 15th, and also building the corresponding binary pattern. Hence, the binary informed patterns proposed are
Pattern 1: $\overline{A} = [1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0]$
Pattern 2: $\overline{A} = [1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0]$.

This is certainly an oversimplification of the real rhythmic patterns, since it does not take into account the accented and muffled strokes that are an essential trait of a *piano* performance. It would be possible to encompass dynamic variations into the informed pattern by considering distinct quantized values of the feature for different type of strokes. However, the binary patterns were favoured for the sake of simplicity and as a proof of concept.

Table 1 [5] compares 4 general-purpose beat-tracking algorithms with the proposed algorithm using the binary informed patterns, and also (for conciseness) the experiments discussed in the next section. Results are averaged over the whole database and weighted by the number of beats and downbeats of each audio file. Although the beat rate (or a multiple) is sometimes precisely estimated by the general-purpose beat-tracking algorithms, the correct metrical level and/or the phase of the beat is usually misidentified.

	BEAT			DOWNBEAT	
	CMLt	AMLt	Fmea	CMLt	Fmea
General-purpose					
Ellis [7]	44.2	63.0	43.8	–	–
Dixon [6]	13.9	14.9	22.7	–	–
IBT [19]	9.1	27.6	16.7	–	–
Klapuri [13]	28.8	35.5	29.3	36.6	13.2
Informed patterns – Section 4.3					
Pattern 1	80.2	80.5	81.3	84.7	79.1
Pattern 2	79.0	81.0	79.8	81.2	77.5
Learned patterns – Section 4.4 (leave-one-out)					
Median	79.9	79.9	80.8	82.4	76.9
K-means 2	81.7	81.7	82.6	84.4	79.3
K-means 5	82.5	82.5	83.6	85.2	80.6

Table 1. Performance of the different algorithms considered.

4.4 Experiments with learned rhythmic patterns

The labeled database allows the study of the rhythmic patterns actually present in real performances. There are different possible approaches to extract a single rhythmic pattern to track from the annotated data. For each *tatum*-grid position in the bar-length pattern, all the feature values in the dataset can be collected, and their distribution can be modeled, e.g. by a GMM as in [14]. The distribution of feature values in the low-frequency range will be dominated by the *base* patterns of the *piano* drum, albeit there will be a considerable amount of *repicado* patterns [21]. In order to cope with that, a simple model was chosen: the median of feature values for each *tatum* beat, which is less influenced by outliers than the mean.

The problem with the median pattern is that it models different beat positions independently. A better suited approach is to group the patterns based on their similarity into a given number of clusters, and select the centroid of the majority cluster as a good prototype of the *base* pattern. This was applied in [21] to identify *base* patterns

[4] Computed with standard settings using code at https://code.soundsoftware.ac.uk/projects/beat-evaluation/.

[5] Additional details can be found in http://www.eumus.edu.uy/candombe/papers/ISMIR2015/.

of the *piano* drum in a performance, and similarly in [10] to learn rhythmic patterns from annotated data to adapt a beat-tracking model to specific music styles. Figure 4 shows the patterns learned from the whole database, using the median and the centroid of the majority cluster obtained with K-means for 2 and 5 clusters. It is remarkable that the differently learned patterns are quite similar, exhibiting the syncopated 4th *tatum* beat as the most accented one. The locations of articulated beats for the informed patterns of the previous section are also depicted, and are consistent with the learned patterns. The K-means approach turned out to be little sensitive to the number of clusters, yielding similar patterns from 1 to 6.

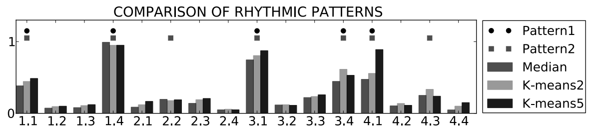

Figure 4. Comparison of the different patterns considered. (Median and K-means learned from the whole database.)

For testing the performance of the learning approach a leave-one-out scheme was implemented and the results are detailed in Table 1. Not surprisingly, performance is almost the same for the different rhythmic patterns. Considering different feature values instead of binary patterns did not yield any notable performance increase.

A detailed inspection of the performance attained for each recording in the database, as depicted in Figure 5, shows there is still some room for improvement, given that about half-a-dozen files are definitely mistracked. This may indicate that the pattern $\overline{A}$ to track simply does not properly match the given performance. To check this hypothesis, a K-means (K=2) clustering was carried out only with the candidate patterns found within each target recording, whose tracking was then performed using the centroid of the majority cluster as $\overline{A}$. Table 2 shows the new results obtained for the files with lower performance (CMLt$<$50%) in the dataset. Except for the first one, performance was (sometimes notably) improved when the informed rhythmic pattern is the one that better matches the recording. Therefore, modeling several rhythmic patterns as in [10] can potentially improve the current results.

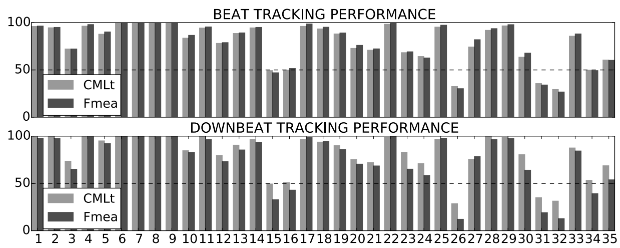

Figure 5. Leave-one-out performance for each recording of the database using the K-means pattern with K=2.

	BEAT		DOWNBEAT	
Recording #	CMLt	Fmea	CMLt	Fmea
15	34.1	32.8	32.7	7.2
16	95.6	98.0	96.3	97.1
26	40.2	36.9	42.9	22.2
31	71.3	69.9	78.3	67.8
32	55.7	54.1	59.6	44.7
34	60.9	60.0	62.7	51.7

Table 2. Scores attained when tracking the centroid of the majority cluster for each of the low performing files.

5. DISCUSSION AND FUTURE WORK

This paper tackled the problem of automatic rhythmic analysis of Candombe audio signals. A study of the rhythmic structure of Candombe was described, along with a pattern tracking algorithm that could deal with the particular characteristics of this rhythm. From the rhythm description and the presented experiments, it becomes clear that typical assumptions of general-purpose beat-tracking algorithms (such as strong events at beat times) do not hold, which hinders their performance. In order to overcome this problem, the proposed algorithm tracks a rhythmic pattern that informs when a beat with or without accentuation is expected to occur, which eventually can determine the complete metric structure. Indeed, experiments employing both rhythmic patterns based on musical knowledge and others learned from a labeled database, showed that the proposed algorithm can estimate the beat and downbeat positions for Candombe whereas traditional methods fail at these tasks. The attained CMLt score of about 80 % for beat tracking is approximately what one can expect from a state-of-the-art algorithm in a standard dataset [2, 9], and what is reported in [10] for a Bayesian approach adapted to a culturally diverse music corpus. The present work gives additional evidence of the generalizability of the Bayesian approach to complex rhythms from different music traditions. The analysis of examples with low performance scores indicates that tracking several rhythmic patterns simultaneously, as proposed in [10], is a promising alternative for future work. Surely taking into account the timbre characteristics of different drums can be profitable.

Along with the annotated database employed, a software implementation of the proposal is being released with the publication of this paper to foster reproducible research (the first available implementation of the Bayesian approach for beat tracking, to the best of our knowledge). [6]

6. ACKNOWLEDGMENTS

This work was supported by funding agencies CAPES and CNPq from Brazil, and ANII and CSIC from Uruguay.

[6] Available from `github.com/lonnes/RhythmicAnalysis`.

7. REFERENCES

[1] G. Andrews. *Blackness in the White Nation: A History of Afro-Uruguay*. The University of North Carolina Press, Chapel Hill, USA, 2010.

[2] S. Böck, F. Krebs, and G. Widmer. A multi-model approach to beat tracking considering heterogeneous music styles. In *Proc. of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 603–608, Taipei, Taiwan, Oct. 2014.

[3] M. Davies and S. Böck. Evaluating the evaluation measures for beat tracking. In *Proc. of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 637–642, Taipei, Taiwan, Oct. 2014.

[4] M. Davies, N. Degara, and M. Plumbley. Evaluation methods for musical audio beat tracking algorithms. Technical Report C4DM-TR-09-06, Center for Digital Music, Queen Mary University, London, UK, Oct. 2009.

[5] N. Degara, E. A. Rua, A. Pena, S. Torres-Guijarro, M. E. P. Davies, and M. D. Plumbley. Reliability-informed beat tracking of musical signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):290–301, Jan. 2012.

[6] S. Dixon. Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58, Aug. 2001.

[7] D. P. W. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, May 2007.

[8] L. Ferreira. An afrocentric approach to musical performance in the black south atlantic: The candombe drumming in Uruguay. *TRANS-Transcultural Music Review*, 11:1–15, Jul. 2007.

[9] A. Holzapfel, M. E. P. Davies, J. R. Zapata, J. L. Oliveira, and F. Gouyon. Selective sampling for beat tracking evaluation. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(9):2539–2548, Nov. 2012.

[10] A. Holzapfel, F. Krebs, and A. Rinivasamurthy. Tracking the "odd": Meter inference in a culturally diverse music corpus. In *Proc. of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*, pages 425–430, Taipei, Taiwan, Oct. 2014.

[11] T. Jehan. *Creating Music by Listening*. PhD thesis, Program in Media Arts and Sciences, Massachusetts Institute of Technology, Cambridge, USA, Sep. 2005.

[12] L. Jure. Principios generativos del toque de repique del candombe. In C. Aharonián, editor, *La música entre África y América*, pages 263–291. Centro Nacional de Documentación Musical Lauro Ayestarán, Montevideo, 2013. In Spanish.

[13] A. P. Klapuri, A. J. Eronen, and J. T. Astola. Analysis of the meter of acoustic musical signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1):342–355, Jan. 2006.

[14] F. Krebs, S. Bck, and G. Widmer. Rhythmic pattern modeling for beat and downbeat tracking in musical audio. In *Proc. of the 14th International Conference on Music Information Retrieval (ISMIR 2013)*, pages 227–232, Curitiba, Brazil, Nov. 2013.

[15] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, Cambridge, USA, 1985.

[16] J. London. Cognitive constraints on metric systems: Some observations and hypotheses. *Music Perception*, 19(4):529–550, Summer 2002.

[17] M. F. McKinney and D. Moelants. Deviations from the resonance theory of tempo induction. In *Proc. of the Conference on Interdisciplinary Musicology (CIM04)*, pages 1–11, Graz, Austria, Apr. 2004.

[18] K. P. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, Computer Science Division, University of California, Berkeley, Berkeley, USA, Fall 2002.

[19] J. L. Oliveira, M. E. P. Davies, F. Gouyon, and L. P. Reis. Beat tracking for multiple applications: A multi-agent system architecture with state recovery. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(10):2696–2706, Dec. 2012.

[20] Geoffroy Peeters. Template-based estimation of time-varying tempo. *EURASIP Journal on Advances in Signal Processing*, 2007(67215):1–14, Dec. 2006.

[21] M. Rocamora, L. Jure, and L. W. P. Biscainho. Tools for detection and classification of piano drum patterns from candombe recordings. In *Proc. of the 9th Conference on Interdisciplinary Musicology (CIM14)*, pages 382–387, Berlin, Germany, Dec. 2014.

[22] A. Srinivasamurthy, A. Holzapfel, and X. Serra. In search of automatic rhythm analysis methods for turkish and indian art music. *Journal of New Music Research*, 43:94–114, Mar. 2014.

[23] D. Temperley. *The Cognition of Basic Musical Structures*. MIT Press, Cambridge, USA, 2001.

[24] N. Whiteley, A. Cemgil, and S. Godsill. Bayesian modelling of temporal structure in musical audio. In *Proc. of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*, pages 29–34, Victoria, Canada, Oct. 2006.

[25] M. Wright, W. A. Schloss, and G. Tzanetakis. Analyzing afro-cuban rhythms using rotation-aware clave template matching with dynamic programming. In *Proc. of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 647–652, Philadelphia, USA, Sep. 2008.