Sociedad de Ingeniería de Audio
Artículo de Congreso
Congreso Latinoamericano de la AES 2018
24 a 26 de Septiembre de 2018
Montevideo, Uruguay

Este artículo es una reproducción del original final entregado por el autor, sin ediciones, correcciones o consideraciones realizadas por el comité técnico. La AES Latinoamérica no se responsabiliza por el contenido. Otros artículos pueden ser adquiridos a través de la Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Información sobre la sección Latinoamericana puede obtenerse en www.americalatina.aes.org. Todos los derechos son reservados. No se permite la reproducción total o parcial de este artículo sin autorización expresa de la AES Latinoamérica.

A Novel Dataset of Brazilian Rhythmic Instruments and Some Experiments in Computational Rhythm Analysis

Lucas S. Maia,[1] Pedro D. de Tomaz Júnior,[2] Magdalena Fuentes,[3,4] Martín Rocamora,[5] Luiz W. P. Biscainho,[1] Maurício V. M. da Costa,[1] and Sara Cohen[6]

[1] Federal Univ. of Rio de Janeiro, PEE/COPPE; Rio de Janeiro, RJ, 21941-972, Brazil

[2] Federal Univ. of Amazonas, Faculty of Technology; Manaus, AM, 69077-000, Brazil

[3] L2S, CNRS-Univ.Paris-Sud-CentraleSupélec; Gif-sur-Yvette, France

[4] LTCI, Télécom ParisTech, Univ. Paris-Saclay; Paris, France

[5] Universidad de la República, Instituto de Ingeniería Eléctrica; Montevideo, 11300, Uruguay

[6] Federal Univ. of Rio de Janeiro, DMEM/EM; Rio de Janeiro, RJ, 20021-290, Brazil

lucas.maia@smt.ufrj.br, pedrodonadio7@gmail.com, magdalena.fuentes@l2s.centralesupelec.fr, rocamora@fing.edu.uy, wagner@smt.ufrj.br, mauricio.costa@smt.ufrj.br, saracohen.ufrj@gmail.com

ABSTRACT

This paper describes and releases the Brazilian Rhythmic Instruments Dataset, assembled for research in Music Information Retrieval. It is comprised of 274 solo- and 93 multiple-instrument recorded tracks of 10 different instrument classes with variations (e.g., material, size, stick) playing in 5 main rhythm classes (*samba*, *partido alto*, *samba-enredo*, *capoeira*, and *marcha*). We describe the dataset recording process and content, and a few examples of challenges it launches for computational rhythm analysis.

0 INTRODUCTION

Within the Music Information Retrieval (MIR) community, there are several music audio databases available for tasks such as genre classification, tempo and beat tracking, chord recognition and music structure analysis [1]. These corpora are composed of either synthetic or real audio items, which in turn might have been either recorded for research purposes or sampled from commercial recordings [2]. Most available databases are composed of Western music, thus a large portion of MIR research is focused on music genres such as pop, rock, jazz or classical. Care must be taken when applying the methods developed for this kind of music to non-Western music collections [3, 4]. Furthermore, when present, non-Western music is generally clustered into large heterogeneous labels such as "World" or "Latin". For instance, in the Ballroom Dataset [5], we verified that 14 out of the 86 excerpts labeled as *samba* could actually be considered as from Brazilian roots. Only half of those 14 excerpts represent *samba* sub-genres (e.g., *bossa nova*, *pagode* etc.), while the other half could actually be classified as *axé*, a totally different Brazilian rhythm. With a similar trend spotted in the Extended Ballroom Dataset [6], it is clear that the dancer's "ballroom *samba*" is very different from Brazilian *samba*, which has to be taken into consideration for computational musicology research.

In this paper we present a novel dataset of recorded Brazilian rhythmic instruments. The Brazilian Rhythmic Instruments Dataset (BRID) is copyright-free and contains short solo- and multiple-instrument tracks in different Brazilian rhythmic styles. As a form of exemplifying the need for (and importance of) such a resource, this work includes a selected group of experiments on beat and downbeat tracking, rhythmic pattern recognition, and microtiming analysis performed on *samba* samples from the BRID. Besides illustrating the richness of musicological content embedded in such material, as a by-product the examples highlight how some specificities of Brazilian rhythms may mislead state-of-the-art methods for computational analysis originally tailored for traditional Western music, and thus call for the development of more robust MIR tools.

The paper is organized as follows. In Section 1, we describe the process of recording the dataset as well as the instruments and rhythms selected to be part of it. Section 2 illustrates how the dataset can be used in MIR research through a set of experiments on computational rhythmic analysis. The last section draws concluding remarks and suggestions for future work.

1 DATASET

The BRID was originally developed in the context of sound source separation [7], but its applicability can be extended to other areas, computational rhythm analysis in particular. The dataset contains 367 short tracks of around 30 s on average, totaling 2 hrs 57min. The tracks consist of recordings of solo or multiple instruments, playing characteristic Brazilian rhythms.

1.1 Instruments and Rhythms

The recorded instruments were selected among the most representative ones in Brazilian music, more specifically in *samba* music. Ten different instrument classes were chosen: *agogô*, *caixa* (snare drum), *cuíca*, *pandeiro* (frame drum), *reco-reco*, *repique*, shaker, *surdo*, *tamborim* and *tantã*. To provide a variety of sounds, both membranophones and idiophones were featured. Also, whenever possible, instruments were varied in shape, size, material (e.g., leather or synthetic drumhead), pitch/tuning (e.g., in a *samba school*,[1] *surdos* are usually tuned in three different pitch ranges) and in the way they were struck (e.g., with the hand, or with a wooden or a plastic stick), spanning a total of 32 variations. For example, the dataset features two *caixa* variations (12" in diameter with either 4 or 6 snare wires), six *pandeiro* variations (either 10", 11" or 12" in diameter with a leather or nylon drumhead) and three *tamborim* variations (one with a leather head struck with a wooden stick, and another one with a nylon head struck with either a wooden or a plastic stick[2]). Figure 1 shows the instrument classes considered.

Table 1: Tempi/number of solo tracks per rhythm.

Rhythm	Tempo (bpm)	*# Tracks*
Samba (SA)	80	54
Partido alto (PA)	100	55
Samba-enredo (SE)	130	60
Marcha (MA)	120	27
Capoeira (CA)	65	12
Samba - virada (VSA)	75 or 80	3
Partido alto - virada (VPA)	75 or 100	36
Samba-enredo - virada (VSE)	130	17
Marcha - virada (VMA)	120	8
Other (OT)	-	2

Table 2: Number of multi-instrument tracks per rhythm.

Rhythm	*# Tracks*
Samba (SA)	41
Partido alto (PA)	28
Samba-enredo (SE)	21
Marcha (MA)	3
Capoeira (CA)	-

The recordings present instruments played in different Brazilian rhythmic styles. Although *samba* and two sub-genres (*samba-enredo* and *partido alto*) have been favored, BRID also features *marcha*, *capoeira*, and a few tracks of *baião* and *maxixe* styles. The number of tracks per rhythm is summarized in Tables 1 and 2.

All featured rhythms are in duple meter. *Samba* is specially known for this type of bar division and for the accentuation of the second beat [8]. Only combinations of instruments and rhythms that are traditionally seen in Brazilian music were considered, to provide a faithful representation of each rhythm.

1.2 Dataset Recording

All the recordings were made in a professional recording studio in Manaus, Brazil, between October

[1] A popular association for the practice of *samba*. *Samba schools* are usually strongly connected to a specific community, where their social events take place and to whom they provide several social services. The climactic event for *samba schools* is the annual carnival parade, when imbued with communal effort they compete for the title.

[2] A leather-head *tamborim* is not played with a plastic drum stick.

(a) *Agogô* (AG) (b) *Caixa* (CX) (c) *Cuíca* (CU) (d) *Pandeiro* (PD) (e) *Reco-reco* (RR)

(f) *Repique* (RP) (g) Shaker (SK) (h) *Surdo* (SU) (i) *Tamborim* (TB) (j) *Tantã* (TT)

Figure 1: Instrument classes.

and December of 2015. The recording room has rectangular shape with dimensions of 4.3 m $\times$ 3.4 m $\times$ 2.3 m and is acoustically treated with a combination of wood and acoustic foam.

Both microphone model and positioning were optimized to translate the sound of each instrument as naturally as possible in the recording, considering the instrument size and the room acoustics. Most instruments were recorded with dynamic microphones within a distance of around 20 cm. The digital files were recorded with a sampling rate of 44.1 kHz and 16-bit resolution.

There are two groups of tracks in the dataset. The first one consists of instruments recorded solo, with the musicians performing in various Brazilian styles following a metronome track. Three musicians were recorded separately, each playing around 90 different instrument–rhythm combinations. For each instrument class, there is at least one track that consists of a *virada* of one of the main rhythms.[3] These are free improvisation patterns (still subject to the metronome track), which are very common in *rodas de samba*.[4] It is worth mentioning that the musicians brought their own instruments for the recording sessions. Although the general characteristics of each instrument are the same, e.g., size and type of material, subtle differences in construction bring additional timbre variability to the dataset.

The second set of tracks of the dataset gathers group performances, with the musicians playing together different rhythmic styles without a metronome reference. The instruments were individually captured with directional microphones, which were strategically positioned to minimize sound bleed, and two condenser microphones in omni polar pattern captured the overall sound in the room. The performances were designed to emulate typical arrangements of each style. Following this procedure, 19 recordings were made with four musicians, 29 with three musicians, and 45 with two musicians playing at a time.

[3]Except for shaker tracks.

[4]A small and informal gathering to play and dance to *samba* music. It is a communal practice highly characterized by improvisation where musicians and dancers interact and learn with one another.

1.3 Track Labeling

Each audio track is given a unique filename, which starts with a four-digit number between brackets—a global identification number `[GID#]`, sequential for the entire dataset.

In solo track (`S`) filenames, the `GID#` is followed by four groups of characters, whose format is either `SW-XXX-YY-ZZ` or `SW-XXX-YY-VZZ`, where `W` is the number for the musician playing in the track, `XXX` specifies the instrument class and variation being played, `YY` consists of a counter for tracks with the same pair musician–instrument, and `ZZ` (or `VZZ`) indicates the rhythmic style (or a *virada* for that style).

For acoustic mixture tracks (`M`), the `GID#` is followed by three groups of characters, whose format is either `MW-YY-ZZ` or `MW-YY-VZZ`. Here, `W` indicates the number of instruments recorded in the track, `YY` is the counter for a given `MX` prefix, and `ZZ` (or `VZZ`) means the same as in the case of solo tracks. The unique identifier for each instrument class and for each rhythm can be found in Figure 1, and in Tables 1 and 2, respectively.

Two samples from the dataset: file `[0192] S2-PD3-01-SA.wav` contains a solo recording of a *pandeiro* (variation 3: 11"; leather head) being played by musician #2 in a *samba* style; and file `[0010] M4-10-SE.wav` is a *samba-enredo* track performed by four musicians.

2 EXPERIMENTS

Some experiments on computational analysis of musical rhythm are discussed in this section in order to show the usefulness of the dataset. First, beat and downbeat tracking tasks are addressed using a representative set of selected audio files exhibiting different characteristics. The experiments show some challenges inherent in *samba* music that arise when tackling these tasks. Secondly, based on the beat and downbeat annotations, a methodology is applied to study the different rhythmic patterns present in a given audio recording.

2.1 Beat and Downbeat Tracking

Musical rhythm involves patterns of organized durations that are phenomenally present in the auditory stream [9], i.e., they form the music stimulus itself. According to modern music theory, the metrical structure can be regarded as a regular pattern of points in time, hierarchically organized into levels [10]. However, these points do not necessarily correspond to the actual events present in the musical piece. The metrical structure is inferred by the listener. The *beats* specifically refer to the pulsation of the perceptually most salient metrical level (usually the one a person would tap or clap along with the music), which are then further grouped into bars. The first beat of each bar is called the *downbeat*.

Automatic beat and downbeat tracking have received considerable attention in MIR in recent years, as they provide key information for several applications such as automatic music transcription [11], structural segmentation [12], computational musicology [13] or measuring of rhythm similarity [14]. Particularly, the introduction of deep neural networks meant a major improvement in the accuracy of beat and downbeat tracking systems [15, 16, 17], becoming the state of the art. Such approaches usually consist of a first stage of high-level feature extraction with neural networks, whose outcome is an activation function that indicates the most likely candidates for beats and/or downbeats. This is often followed by a post-processing stage that corresponds to a dynamic model such as Hidden Markov Models, Dynamic Bayesian Networks (DBNs) or Conditional Random Fields [18, 19, 20]. This temporal models are commonly constrained in order to account for high level information such as common or impossible sequences of events, and to use that information to infer the metrical structure of the musical piece.

Three state-of-the-art systems for beat and downbeat tracking are adopted in this work, namely MMB14 [16], JB16 [18], and K16 [17], which are briefly described in the following. We use the implementations available in the `madmom` package.[5][6] Böck et al. [16], proposed a system for beat tracking that uses multiple recurrent neural networks which are specialized in certain music styles. Each recurrent network consist of a concatenation of three Bi-Directional Long-Short Term Memory (Bi-LSTMs) hidden layers with 25 units per layer. The system chooses the most appropriate beat activation function for the given input signal by comparing the respective activation functions with a reference network trained in all music styles. Finally, tempo and beat phase are determined using a DBN. Then, Böck et al. [18], presented a system that jointly tracks beats and downbeats using Bi-LSTMs. The authors used three different magnitude spectrograms and their first order difference as input representations, in order to help the networks capture features precisely in both time and frequency. These representations were fed into a cascade of three fully-connected Bi-LSTMs, obtaining activation functions for beat and downbeat as output. Subsequently, a highly constrained DBN was used for inferring the metrical structure. In turn, Krebs et al. [17] proposed a downbeat tracking system that uses two beat-synchronous features, representing the percussive and harmonic content of the audio signal. These representations, based in spectral flux and chroma, are then fed into two independent Bi-Directional Gated Recurrent Units, whose output is averaged to obtain the downbeat likelihood. The inference over downbeat candidates relies on a constrained DBN.

[5]Madmom package version 0.16 [21].

[6]Note that only JB16 tracks both beats and downbeats, while MMB14 performs beat tracking, and K16 does downbeat tracking.

Table 3: Selected *solos*.

Filename	Instrument
`[0218] S2-TB3-01-SE`	*Tamborim*
`[0229] S2-CX2-02-PA`	*Caixa*
`[0258] S2-SK2-02-PA`	Shaker
`[0280] S2-SU2-05-SE`	*Surdo*

Table 4: Selected *mixtures*.

Filename	Instruments
`[0013] M4-13-SE`	*Cuíca, caixa, tamborim, surdo*
`[0039] M3-20-SE`	*Caixa, tamborim, tantã*
`[0047] M3-28-SE`	*Caixa, surdo, surdo*
`[0051] M2-03-SA`	*Tantã, surdo*

A set of 8 audio files, representative of the content of the dataset, was selected for the experiments. It comprises 4 *solos* and 4 *mixtures*, involving different rhythms (*samba*, *samba-enredo* and *partido alto*) and different instruments, as shown in Tables 3 and 4.

We report the F-measure[7] score (F), a commonly used metric that is computed from correctly detected beats (or downbeats) within a window of 70 ms by:

$$F = \frac{2PR}{P+R},$$

where P (Precision) denotes the ratio between correctly detected beats and all detected beats, and R (Recall) denotes the ratio between correctly detected beats and the total number of annotated beats.

2.1.1 Analysis of the selected *solos*

Many instruments, such as *tamborim* or *shaker*, have a rhythmic pattern that repeats itself within each beat (see Figure 2). Hence, it is very difficult (even for an expert) to establish the location of downbeats from a

[7]Considering that the evaluation of beat tracking algorithms was not the focus of this paper, other evaluation metrics were not used.

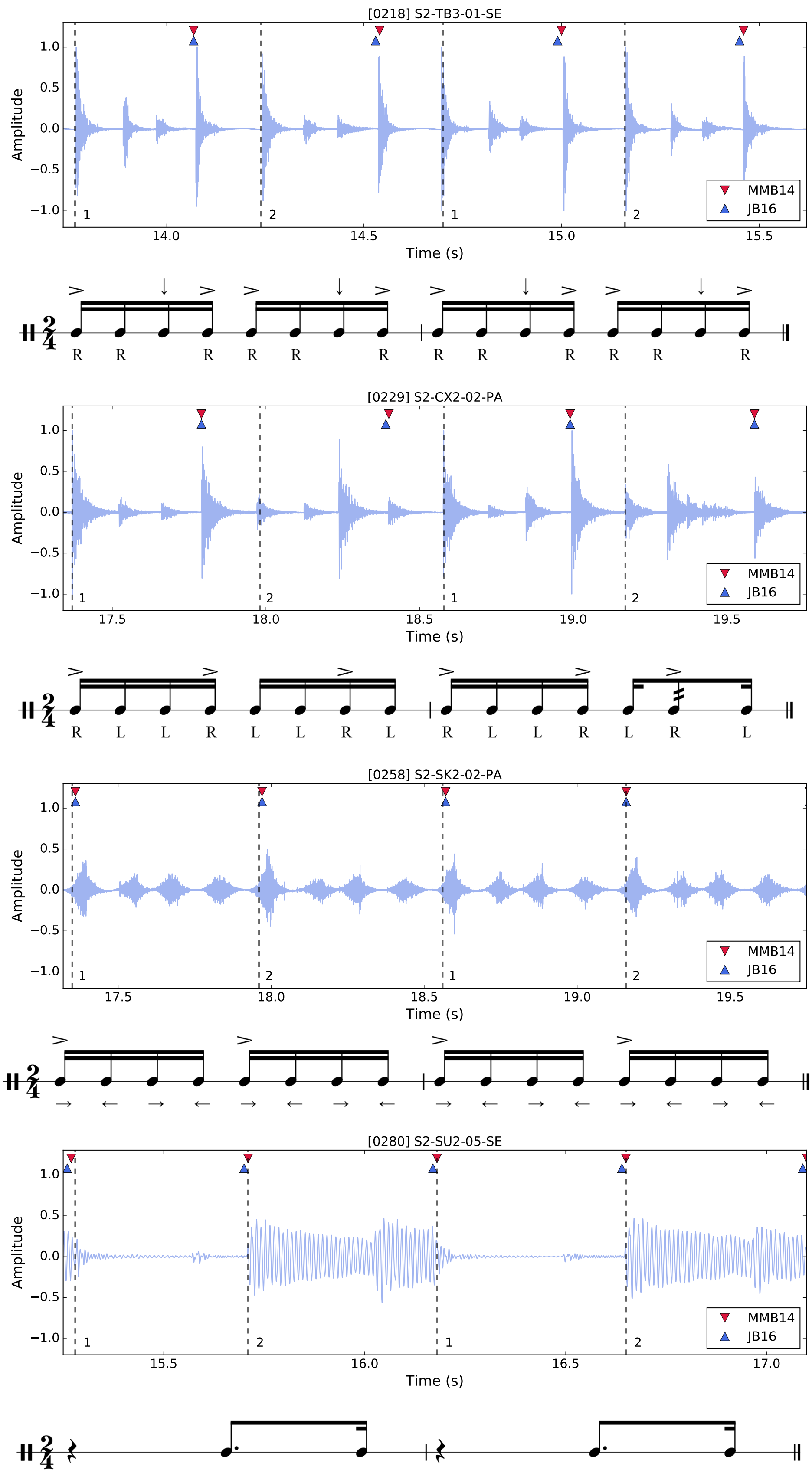

Figure 2: Beat tracking for the selected *solo* track examples. From top to bottom: *tamborim*, *caixa*, shaker and *surdo*. The waveform plots show two bars of the rhythmic patterns, with vertical lines indicating annotated beats. The estimated beats are depicted with markers. Rhythmic patterns are schematically represented in music notation.

solo track without any further references. For this reason, only beat tracking is tackled in this section.

The beat positions for each *solo* track are estimated using the MMB14 [16] and JB16 [18] algorithms. The obtained results are presented in Table 5. A two-bar length excerpt of each audio file is shown in Figure 2, which depicts the annotated beat positions and the beat estimations for each algorithm. The annotations also indicate beat number in a 2/4 meter (i.e. 1 and 2). The rhythmic patterns in music notation are also provided.

The performance of both algorithms is very similar: they miss the phase of the beat in two of the files (0218 and 0229) and correctly track the other two (0258 and 0280). A detailed inspection of Figure 2 makes it clear that the troublesome rhythmic patterns, i.e. those of the *tamborim* and *caixa*, have strong phenomenological accents displaced with respect to the metric structure. Conversely, the pattern of the shaker accentuates every beat. In the case of the *surdo*, there are actually several different rhythmic patterns played, but most of the time the second beat of the bar is strongly articulated. This distinctive trait of *samba* rhythm, while advantageous for beat tracking, proved to be very challenging for downbeat estimation, as shown next.

2.1.2 Analysis of the selected *mixtures*

As for the mixtures, both beats and downbeats are tracked. The beats are estimated using the MMB14 [16] and JB16 [18], while the downbeats are estimated with K16 [17] and JB16 [18]. Since all the mixtures are in 2/4, we set the search-space of the DBN for downbeat tracking to bar lengths of 2 and 4 beats, both yielding the same results. Tables 6 and 7 show the beat and downbeat tracking results, respectively.

Whereas beat tracking in the mixtures is not problematic (except for file 0051, in which half of the estimates are out of phase probably due to the anacrusis at the beginning), downbeat tracking is very challenging. Both algorithms fail to correctly track the downbeats for all the recordings. The downbeat estimates tend to follow the second beat, suggesting that the *surdo* is misleading the algorithms, as shown in Figure 3.

Table 5: Beat F-measure scores for solos.

Model \ Track	0218	0229	0258	0280
MMB14 [16]	0.00	0.00	1.00	0.96
JB16 [18]	0.00	0.00	1.00	1.00

Table 6: Beat F-measure scores for mixtures.

Model \ Track	0013	0039	0047	0051
MMB14 [16]	0.99	0.98	0.98	0.56
JB16 [18]	1.00	1.00	1.00	0.40

Table 7: Downbeat F-measure scores for mixtures.

Model \ Track	0013	0039	0047	0051
K16 [17]	0.00	0.00	0.00	0.00
JB16 [18]	0.00	0.00	0.00	0.00

2.2 Pattern Extraction and Microtiming

In this experiment, we combine the knowledge of beat and downbeat locations (automatically annotated and manually corrected) with an onset detection function [22] in order to study the rhythmic patterns present in a given audio track, following the approach presented by Rocamora et al. in [23]. Assuming that in *samba* the lowest pulsation level can usually be found at the 16th-note level, from downbeat locations we can estimate an isochronous grid at this metrical level. The peaks of the onset function which are closer to each point in this assembled grid are considered to have been planned at those metrical positions. Deviations, which characterize microtiming expression, can be thought as a result of both physical/technical limitations and of implicit stylistic/personal requirements (“swing”).

The algorithm is as follows. Firstly, the spectral flux is computed from the Short-Time Fourier Transform of the signal (in the MEL scale), for 40-ms windows with hop size of 2 ms. This onset function is then normalized by the 8-norm in a sliding window of length equal to half of the average bar length (estimated from beat/downbeat data). Finally, the normalized onset function is time-quantized considering an isochronous grid anchored at downbeat pulses. The maximum value in a 100 ms window centered at each frame of the isochronous grid is taken as a feature value, yielding an 8-dimension feature vector per bar. Vectors from adjacent bars can be grouped as to form larger rhythmic patterns—here called cycles (e.g., a 16-dimension vector for 2-bar patterns). In each feature vector, articulated grid instants are represented by a high value (close to 1.0) at the corresponding dimension. The feature vectors are clustered by an unsupervised clustering algorithm, e.g., K-means, to aid in the analysis of the different kinds of patterns present in the recording—cluster centroids act as templates for the patterns. The feature vectors can then be displayed as rows of a cycle-length feature map [23], in which the darker the value of a cell, the higher its value in the respective dimension. One such map can be seen in Figure 4 for an *agogô* recording in *samba-enredo* rhythm (file 0251). In this map, there are two different 2-bar length patterns that can be readily identified and that agree with the analysis of the recording by an expert musician. It is possible to see that, in the case of *samba* music, strong pulses are not always found at the start of a bar. In fact, for each rhythmic cycle (feature vector) in this example, the second beat of both bars, the first beat of the second bar and the last “tatum”[8] are the ones competing for the role of strongest pulse.

Microtiming properties can also be studied by analyzing, for example, deviations of the articulated notes from their expected positions in the isochronous grid at each cycle. Figure 5 shows the calculated deviations (in percentages, relative to the average tatum period) for each point in the isochronous grid for the two patterns

[8] In the MIR community, the tatum is usually defined as the lowest level of the metrical hierarchy.

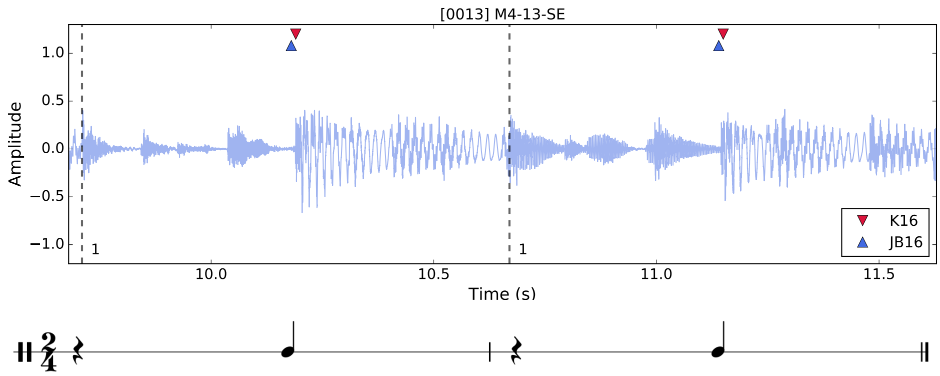

Figure 3: Downbeat tracking for one of the selected *mixture* track examples. The waveform plot shows two bars, with vertical lines indicating the annotated downbeats. The estimated downbeats are indicated by markers. The music notation shows the *surdo* onsets at the second beat of each bar, which is troublesome for downbeat detection.

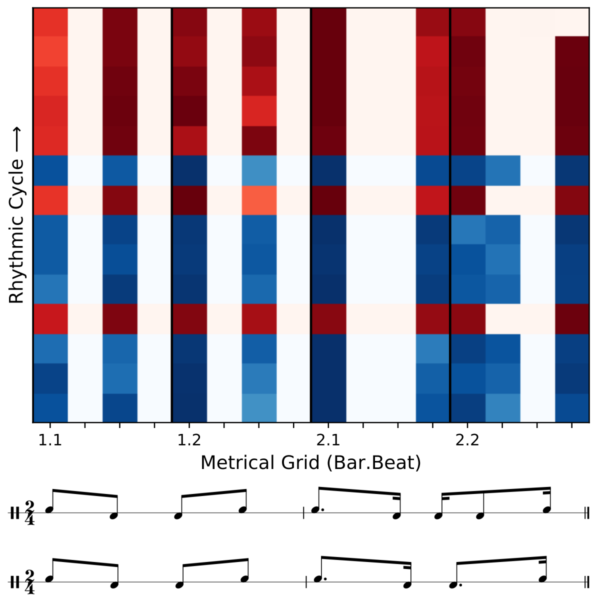

Figure 4: Feature map for the *agogô* recording. Two patterns (blue and red, transcribed as lines 1 and 2 above, respectively) are evident from this feature map.

found in the *agogô* recording. As downbeats determine the grid, no deviation is shown at these points. On average, the second beat of the first bar falls on the grid, whereas at the second bar it is slightly delayed. This is probably due to the patterns themselves, in which the musician has to strike two or three notes in rapid succession. All other points in the metrical grid are continuously played ahead of time, with the third tatum of the first bar showing the highest deviation (almost 23 ms ahead of position at a tempo of 130 bpm). Gouyon reported a similar tendency to play ahead of the quantized grid in *samba de roda* recordings in [24].

3 CONCLUSION

As computational musicology advances, research efforts must be made to encompass musical traditions other than Western in its models, in order to transform it into a truly multicultural field. We have described here the Brazilian Rhythmic Instruments Dataset, a

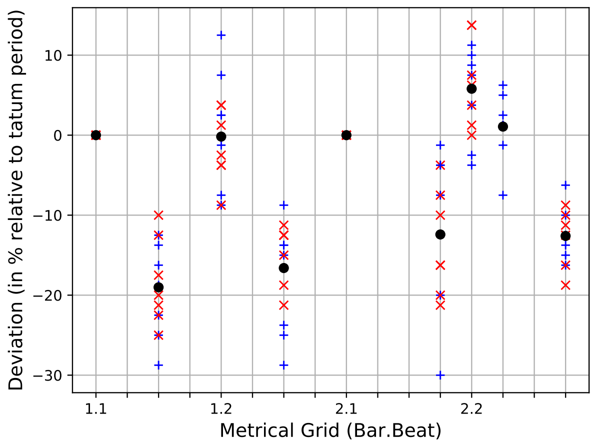

Figure 5: Deviations from the isochronous grid in the *agogô* recording. The different patterns are identified following the color convention of Figure 4.

copyright-free dataset for research within the MIR community. Our experiments have shown that state-of-the-art beat/downbeat tracking algorithms do not perform well when facing specificities of *samba* music. This indicates the need for better conditioning of such models, which is left as future work. The dataset also provides interesting paths in the modeling of microtiming properties for *samba* and its sub-genres. It could also be used in the task of tempo estimation, although state-of-the-art algorithms would probably not find it to be much of a challenge due to the lack of tempo variations. At the time of writing, annotations of beat, downbeat, and type of articulation are being produced. The dataset contents are available at www.smt.ufrj.br/~starel.

4 ACNOWLEDGMENTS

This work is supported by the STIC-AmSud program though the STAREL project 18-STIC-08.

REFERENCES

[1] M. Müller, *Fundamentals of Music Processing*, Springer Verlag, Berlin, Germany, 2015.

[2] G. Peeters and K. Fort, "Towards a (Better) Definition of the Description of Annotated MIR Cor-

pora," in *13th Int. Soc. for Music Information Retrieval Conf. (ISMIR 2012)*, Porto, Portugal, Oct. 2012, pp. 25–30.

[3] T. Lidy, C. N. Silla Jr., O. Cornelis, et al., "On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-Western and ethnic music collections," *Signal Processing*, vol. 90, no. 4, pp. 1032–1048, Apr. 2010.

[4] X. Serra, "A multicultural approach in music information research," in *12th Int. Soc. for Music Information Retrieval Conf. (ISMIR 2011)*, Miami, USA, Oct. 2011, pp. 151–156.

[5] F. Gouyon, S. Dixon, E. Pampalk, and G. Widmer, "Evaluating rhythmic descriptors for musical genre classification," in *25th AES Int. Conf.*, London, UK, June 2004, pp. 196–204.

[6] U. Marchand and G. Peeters, "The Extended Ballroom Dataset," in *Late-Breaking Session of the 17th Int. Soc. for Music Information Retrieval Conf. (ISMIR 2016)*, New York, USA, Aug. 2016.

[7] P. D. de Tomaz Júnior, "Separação automática de instrumentos de percussão brasileira a partir de mistura pré-gravada," M.S. thesis, Federal University of Amazonas, Manaus, Brazil, 2016.

[8] G. Gonçalves and O. Costa, *The Carioca Groove: The Rio de Janeiro's Samba Schools Drum Sections*, Groove, Rio de Janeiro, Brazil, 2000.

[9] J. London, *Hearing in Time: Psychological Aspects of Musical Meter*, Oxford University Press, New York, USA, 2004.

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

[11] M. Mauch and S. Dixon, "Simultaneous Estimation of Chords and Musical Context From Audio," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1280–1289, Aug. 2010.

[12] N. C. Maddage, "Automatic Structure Detection for Popular Music," *IEEE MultiMedia*, vol. 13, no. 1, pp. 65–77, Jan. 2006.

[13] M. Hamanaka, K. Hirata, and S. Tojo, "Musical Structural Analysis Database Based on {GTTM}," in *15th Conf. of the Int. Soc. for Music Information Retrieval (ISMIR 2014)*, Taipei, Taiwan, Oct. 2014, pp. 325–330.

[14] J. Paulus and A. Klapuri, "Measuring the Similarity of Rhythmic Patterns," in *3rd. Int. Conf. on Music Information Retrieval (ISMIR 2002)*, Paris, France, Oct. 2002, pp. 150–156.

[15] S. Durand, J. P. Bello, B. David, and G. Richard, "Downbeat tracking with multiple features and deep neural networks," in *2015 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 409–413.

[16] S. Böck, F. Krebs, and G. Widmer, "A Multimodel Approach to Beat Tracking Considering Heterogeneous Music Styles," in *15th Conf. of the Int. Soc. for Music Information Retrieval (ISMIR 2014)*, Taipei, Taiwan, Oct. 2014, pp. 603–608.

[17] F. Krebs, S. Böck, M. Dorfer, and G. Widmer, "Downbeat Tracking Using Beat Synchronous Features with Recurrent Neural Networks," in *17th Int. Soc. for Music Information Retrieval Conf. (ISMIR 2016)*, New York, USA, Aug. 2016, pp. 129–135.

[18] S. Böck, F. Krebs, and G. Widmer, "Joint Beat and Downbeat Tracking with Recurrent Neural Networks," in *17th Int. Soc. for Music Information Retrieval Conf. (ISMIR 2016)*, New York, USA, Aug. 2016, pp. 255–261.

[19] S. Durand and S. Essid, "Downbeat Detection With Conditional Random Fields And Deep Learned Features," in *17th Int. Soc. for Music Information Retrieval Conf. (ISMIR 2016)*, New York, USA, Aug. 2016, pp. 386–392.

[20] S. Durand, J. P. Bello, B. David, and G. Richard, "Robust Downbeat Tracking Using an Ensemble of Convolutional Networks," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 25, no. 1, pp. 76–89, Jan. 2017.

[21] S. Böck, F. Korzeniowski, J. Schlüter, F. Krebs, and G. Widmer, "madmom: a new Python Audio and Music Signal Processing Library," in *24th ACM Int. Conf. on Multimedia*, Amsterdam, Netherlands, Oct. 2016, pp. 1174–1178.

[22] S. Dixon, "Onset detection revisited," in *9th Int. Conf. on Digital Audio Effects (DAFx-06)*, Montreal, Canada, Sept. 2006, pp. 133–137.

[23] M. Rocamora, L. Jure, and L. W. P. Biscainho, "Tools for detection and classification of piano drum patterns from candombe recordings," in *9th Conf. on Interdisciplinary Musicology (CIM14)*, Berlin, Germany, Dec. 2014, pp. 382–387.

[24] F. Gouyon, "Microtiming in "Samba de Roda" — Preliminary experiments with polyphonic audio," in *11th Brazilian Symposium on Computer Music (SBCM'2007)*, São Paulo, Brazil, Sept. 2007, pp. 197–203.