SAMBASET: A DATASET OF HISTORICAL SAMBA DE ENREDO RECORDINGS FOR COMPUTATIONAL MUSIC ANALYSIS

Lucas S. Maia[1,2], **Magdalena Fuentes**[3,2], **Luiz W. P. Biscainho**[1], **Martín Rocamora**[4], **Slim Essid**[2]
[1] Federal University of Rio de Janeiro, Brazil
[2] LTCI, Télécom Paris, Institut Polytechnique de Paris, France
[3] L2S, CNRS–Université Paris-Sud–CentraleSupélec, France
[4] Universidad de la República, Uruguay
lucas.maia@smt.ufrj.br

ABSTRACT

In the last few years, several datasets have been released to meet the requirements of "hungry" yet promising data-driven approaches in music technology research. Since, for historical reasons, most investigations conducted in the field still revolve around music of the so-called "Western" tradition, the corresponding data, methodology and conclusions carry a strong cultural bias. Music of non-"Western" background, whenever present, is usually underrepresented, poorly labeled, or even mislabeled, the exception being projects that aim at specifically describing such music. In this paper we present SAMBASET, a dataset of Brazilian *samba* music that contains over 40 hours of historical and modern *samba de enredo* commercial recordings. To the best of our knowledge, this is the first dataset of this genre. We describe the collection of metadata (e.g. artist, composer, release date) and outline our semiautomatic approach to the challenging task of annotating beats in this large dataset, which includes the assessment of the performance of state-of-the-art beat tracking algorithms for this specific case. Finally, we present a study on tempo and beat tracking that illustrates SAMBASET's value, and we comment on other tasks for which it could be used.

1. INTRODUCTION

Machine-learning-based systems in music information retrieval (MIR) are becoming increasingly complex to cope with the also expanding number of tasks and the challenges they propose. In turn, estimating the parameters of these large models requires more and better data [29], especially because such data must often be separated into training, test, and validation sets. Although data augmentation can be used to alleviate this bottleneck [29], this kind of strategy is not able to solve the cultural bias still present in existing MIR data, methodologies, and conclusions [38].

Indeed, a great part of the research in this field focuses on musical traditions usually labeled as "Western". This is worrying, since by doing so we risk not being able to fully evaluate and reproduce specific musical properties found in some cultures [38]. Some datasets attempt to be universal and to cover a large number of music styles, but end up sacrificing the very representation of what they are trying to portray. This is the case, for example, of the well-known Ballroom and Extended Ballroom datasets, whose "Samba" class contains a mixture of songs of different origins, of which only a few examples correspond to Brazilian rhythms, specifically identifiable as *bossa-nova*, *pagode*, and others [28]. In other datasets, music from non-"Western" traditions is given generic labels as "Latin", or "World" [28]. This underscores the importance of increasing the efforts towards the study of non-"Western" traditions found throughout the multicultural world we live in.

1.1 Other Culture-Specific Datasets

Here we review some of the existing datasets devoted to non-"Western" music traditions. One of the biggest projects today is CompMusic [38], which focuses on five particular music cultures: Arab-Andalusian, Beijing Opera, Turkish-makam, Hindustani, and Carnatic. Several annotations are provided, including melody (e.g. singer tonics, pitch contours), rhythm and structure (e.g. tala cycles), scores (e.g. for percussion patterns), and lyrics.

There are also some datasets of Latin-American music launched with MIR in mind. For instance, the dataset released in [32] comprises annotated audio recordings of Uruguayan *candombe* drumming, suited for beat/downbeat tracking. Aimed at music genre classification, the Latin Music Database [39] has Brazilian rhythms—*axé*, *forró*, *gaúcha*, *pagode*, and *sertaneja*—and music from other traditions: *bachata*, *bolero*, *merengue*, *salsa*, and *tango*. Closer to the topic of this article, two datasets focus exclusively on Brazilian music: intended for music genre classification, the Brazilian Music Dataset [40] includes *forró*, rock, *repente*, MPB (Brazilian popular music), *brega*, *sertanejo*, and disco; meant for beat/downbeat tracking and rhythmic pattern analysis, the BRID [28] consists of typi-

cal rhythmic patterns of *samba*, *samba de enredo*, *partido-alto* and other styles played on Brazilian instruments.

1.2 Our Contributions

In this paper, we present SAMBASET, the first large dataset of annotated *samba de enredo* recordings. Besides describing the dataset contents and detailing the beat and downbeat annotation process, we highlight one possible (musicological) use of this dataset through a study on recordings' tempo across the years, and briefly discuss how its results agree with expert knowledge about the evolution of *sambas de enredo* over the last few decades. Finally, we draw our concluding remarks and point out other challenges that can be tackled with SAMBASET.

1.3 Notes on *Samba*

Samba plays a special role in Brazil's image overseas. And every year, the country receives millions of tourists for Carnival activities in cities such as Salvador and Rio de Janeiro. Being Brazil's quintessential rhythm, *samba*'s development is closely related to that of Brazil itself.

Samba's roots can be traced back to dance and religious practices from the Afro-Brazilian diaspora [1, 20] and, as Araujo [1] points out, to the accommodation efforts made by people of African descent to maintain their heritage and cultural identity despite slavery and persecution. In many of these cultural practices, participants would form a *roda* (circle) and accompany one or more dancers (positioned at the center of the *roda*) by clapping, singing, and occasionally playing instruments [1, 37]. These traditions gave origin to different cultural manifestations, collectively associated with the term *samba*, [1] for example: *coco*, *samba de roda*, *partido-alto*, *samba de terreiro*, *pagode*, among others. In the post-Abolition period, *samba* overcame prohibition to become Brazil's national rhythm.

In the 1930s, the genre evolved to the rhythmic framework that still defines it today—generally characterized by duple meter (i.e., binary division of the periodically perceived pulsations) and strong syncopation. However, the idea of syncope—momentary contradiction of the prevailing meter or pulse [36]—can only be adequately applied to "Western" music, creating a fundamental problem in music traditions where this disruption of the pulse is the norm, and not the exception. That is why some authors prefer to resort to the concepts of commetricity and contrametricity [37], which indicate respectively when the surface rhythm confirms or contradicts the underlying meter, a terminology more commonly used in African music studies [2, 24, 37]. Therefore, it is more appropriate to say that *samba* presents a strong tendency towards contrametricity.

Later developments in *samba* led to, arguably, the most internationally famous of all its subgenres, the *samba de enredo*. These are *sambas* subject to an *enredo* (plot) composed in the context of an *escola de samba*—popular association for the practice of *samba*, strongly connected to a specific community—, and presented at parades in an organized competition annually held along a so-called Sambadrome during Carnival. At the core of every *escola de samba* lies the *bateria* (percussion ensemble). During a performance, the rhythmic aura of a *bateria* is created by the superposition of several cyclical individual parts, assigned to each multi-piece instrument set, similarly to what is observed in percussion ensemble practices throughout sub-Saharan Africa [1]. The *bateria* sets the mood of *samba*, but recent studies have observed an increase in the average tempo of *bateria* performances, an effect attributable to stricter parading time constraints [12, 22, 34].

[1] Possibly a variation of *semba*, word for a kind of circle dance practice in the Angolan Kimbundu language [37].

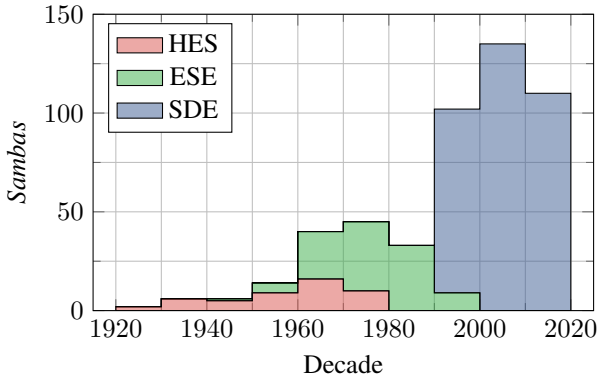


Figure 1: Recordings per decade of first performance.

2. DATASET OVERVIEW

Sambas de enredo are well documented in the phonographic industry. Apart from historical collections, since 1968 the yearly *sambas de enredo* that competing *escolas de samba* will perform at the carnival parade have been professionally recorded and marketed. Initially available as LP records, these official compilations began to appear as CDs in 1990. Since then, the amount of musicians (instrumentalists and choir) in each track has only increased.

Currently comprised of audio recordings, annotations and metadata, SAMBASET covers different eras, from later renditions of old classics to the most recent *sambas de enredo* just out of the Sambadrome. Figure 1 indicates the distribution of *sambas* w.r.t. the year they were first performed (typically, the parading year). Three major collections make up the dataset; in chronological order:

História das Escolas de Samba (HES): a collection of historical *sambas*, composed between 1928 and 1974, from four major *escolas de samba*, arranged and interpreted by the instrumentalists of each *escola*. Recorded in 1974, published in four LPs by Discos Marcus Pereira (redistributed as CDs in 2011), their 48 tracks also include a few *sambas de quadra/de terreiro* and *partidos-altos*.

Escolas de Samba – Enredos (ESE): a collection of historical *sambas*, composed between 1949 and 1993, from ten traditional *escolas de samba* in the voices of many idols from *samba*'s history, accompanied by a selected ensemble of instrumentalists and choir. There are a total of 100 tracks recorded and released in 1993 by Sony Music. This

	Genres			
Escola	SE	ST/SQ	OT	Total
Mangueira	45	3	1	49
Portela	41	5	2	48
Salgueiro	42	5	-	47
Império Serrano	31	5	-	36
Mocidade	35	-	1	36
Beija-Flor	34	1	-	35
Imperatriz	35	-	-	35
Vila Isabel	33	-	-	33
União da Ilha	27	-	-	27
Grande Rio	25	-	-	25
Unidos da Tijuca	24	-	-	24
Viradouro	18	-	-	18
Estácio	16	-	-	16
Porto Da Pedra	15	-	-	15
Caprichosos	12	-	-	12
São Clemente	12	-	-	12
Tradição	11	-	-	11
Other *escolas* (7)	14	-	-	14
Total	470	19	4	**493**

Table 1: Number of recordings in *SAMBASET* separated by *escolas* and by genres: *samba de enredo* (SE), *samba de terreiro/samba de quadra* (ST/SQ), and others (OT). [2]

```
<metadata dataset="SAMBASET"
          curator="John Doe"
          version="0.0.1">
...
  <album title="História das Escolas de Samba – Mangueira"
         arranger="Cartola"
         producer="J. C. Botezelli"
         instrumentalists="Various Artists"
         record_label="Discos Marcus Pereira"
         year_published="2011"
         length="00:29:48"
         total_tracks="12"
         album_code="HES1"
         barcode="7892141643634">
    ...
    <track track_number="6"
           title="Vale Do São Francisco"
           artist="Cartola"
           composer="Cartola and Carlos Cachaça"
           year_recorded="1974"
           year_first_performed="1948"
           genre="samba de enredo"
           length="02:49.226"
           samplerate="44100"
           bpm="78.4"
           start_time="00:07.895"
           end_time="02:49.226"
           checksum="d16974f135f0c374677c0e0db101cfea"/>
    ...
  </album>
...
</metadata>
```

Figure 2: Metadata file excerpt.

collection includes a couple of tracks from different subgenres (*samba de terreiro* and *samba-exaltação*).

Sambas de Enredo (SDE): official compilations of *sambas de enredo* recorded by members of the top *escolas* from Rio de Janeiro, for each carnival parade between 1994 and 2018. The 25 CDs gather 338 tracks, published by RCA/BMG/Sony BMG (1994–2006) and by Universal Music (after 2007), with one *samba de enredo* per track.

Table 1 gives the number of tracks for each *escola de samba* featured in the dataset, by genre. In total, there are 493 recorded sambas in 486 audio tracks, [3] resulting in over 40 hrs 30 min of content. All files are stereo with a sampling rate of 44.1 kHz and 16-bit resolution. Not only the three different collections allow for the coverage of different time periods, but they also have distinct sonorous characteristics. In HES, tracks feature only a few musicians playing very naturally and with great expression, as if they were in a *roda*. For several tracks in the official compilation (SDE), on the other hand, more than fifty instrumentalists play simultaneously while a choir of around the same size accompanies the main singer. Finally, ESE presents smaller ensembles and less expressiveness.

3. METADATA AND ANNOTATIONS

Metadata for albums and tracks were carefully curated and organized in an XML file. The information therein described was primarily obtained from CD booklets and later cross-checked with both the *União Brasileira de Compositores* [4] (lit. Brazilian Union of Composers, UBC) and the *Instituto Memória Musical Brasileira* [5] (Brazilian Musical Memory Institute, IMMuB). Whenever corresponding information was available, data was also checked against online database services such as FreeDB, MusicBrainz or Discogs. Finally, we consulted *samba*-oriented forums and websites for additional, conflicting or missing information.

All XML tags can be seen in Figure 2. While most of these labels are straightforward (e.g. title, composer), some require further clarification. First, the album_code refers to a unique code given to each album in the dataset. Albums from the HES and ESE collections were sequentially numbered, i.e., they are referred to by the codes HES1 to HES4 and ESE1 to ESE10, respectively. For SDE, albums were specified via the publishing year, which is also present in the album's title (i.e., SDE1994–SDE2018). The track_number is used with the album_code to name all audio files (e.g. the metadata in Figure 2 corresponds to file HES1.06.wav). Track's start_time and end_time indicate the time each *samba* starts and ends, respectively. This is invaluable since many *samba de enredo* recordings are preceded by a short introductory speech or song motivating the performance, or succeeded by a "farewell" shout after the music has already stopped. The checksum attributes were filled with the MD5 hash of the track's WAV file, to allow the verification of audio data integrity. Finally, mean bpm values were estimated from the beat annotations described in the following.

As of the writing of this paper, SAMBASET has annotations of beat and downbeat produced according to a semi-

[2] Some imbalance can be observed in the distributions of both genres and *escolas*. However, ST/SQ and OT tracks were only kept for the completeness of the dataset in regard to the CD collections, and the *escolas*' playing styles are not so heterogeneous as to make this imbalance critical.

[3] Some tracks in the ESE collection contain more than one *samba*.

[4] http://www.ubc.org.br/

[5] https://immub.org/

automatic procedure, after the results of the experiment described in Section 4. First, automatically-generated beat annotations were obtained for all audio files using the DBNBeatTracker system, which is available in the `madmom` package [3]—deemed a good candidate for providing reliable beat estimations (cf. Section 4). In a second step, these estimations were checked and manually corrected by one of the authors, who addressed eventual phase errors, and missing/extra beats. Since *samba de enredo* is always in duple meter, downbeats could be manually selected during this second phase. This two-step procedure greatly reduced the amount of manual work necessary to annotate beats and downbeats for this entire dataset.

4. ANALYSIS OF BEAT TRACKERS' PERFORMANCE

In this section, we provide a performance analysis of different state-of-the-art beat tracking systems, which were applied on a subset of short (30-second) excerpts from SAMBASET. Samples were selected according to a criterion based on the mean mutual agreement between beat estimation sequences generated by the algorithms under analysis, inspired by the approach of Holzapfel et al. [21]. This subset was manually annotated using Sonic Visualiser [11] by an expert with an engineering background, knowledgeable of audio technologies, and with many years of experience as a practicing musician, in particular of *samba*. Estimations were evaluated against this ground truth using three different types of metrics.

4.1 State-of-the-art Algorithms Considered

Fourteen algorithms (seen in Table 2) were used for sample selection and performance evaluation. We replaced eight of the algorithms originally featured in Holzapfel et al. [21] with six other algorithms released in following years, most notably those provided in the `madmom` package [3].

The algorithms are implemented in different programming languages and, in a few cases, require different operating systems. We used the Python implementations of AUB (version 0.4.9), ELL (provided by the librosa package [30] version 0.6.2), DEG and MFT (Essentia package [8] version 2.1-beta5-dev), BO1, BO2, and BO3 (madmom package [3] version 0.16.1); and the available releases of DIX (in Java, version 0.5.8), DAV (Vamp plugin in conjunction with the Sonic Annotator [10]), IB1 and IB2 (version 1.0 binaries). Finally, the C++ implementation of KLA was kindly provided by the author.

4.2 Evaluation Measures

Since there is currently no consensus on the evaluation metrics for the beat tracking task [13], with choices depending on the type of application, in this work we adopted the following methods [6] :

F-measure [17]: It is defined as the harmonic mean between precision (ratio between correctly detected and estimated beats) and recall (ratio between correctly detected and annotated beats). Generally, an estimated beat is considered correct if within ± 70 ms around an annotation.

Continuity-based measures [19]: Here, an estimated beat is considered correct if it is within a small tolerance around an annotation, the previous estimation has also been deemed correct, and the inter-beat interval is consistent with the inter-annotation interval within another tolerance—both generally set to 17.5% of the inter-annotation interval. CMLt ("correct metrical level") is the ratio between correct and annotated beats; accepting phase errors of half a beat period or octave errors in estimation yields the AMLt ("allowed metrical level").

Information Gain [14]: Defined as the Kullback-Leibler divergence between the observed beat error histogram (considering the timing errors of all estimated beats within a beat-length window around the annotations) and a uniform one (accounting for a pair of unrelated beat sequences), it spans the range $[0, \log_2(K)]$ bits, where K is the number of bins in the histograms (usually 40).

4.3 Selection of Ground Truth Excerpts

In [21], Holzapfel et al. presented a method for selecting challenging music examples for the beat tracking task without ground truth annotations. To do so, they first calculate the mean mutual agreement (MMA) between sequences estimated by a group of state-of-the-art beat trackers. The mutual agreement between two estimated beat sequences $\{i, j\}$ output by different beat tracking systems is given by the Information Gain in bits:

$$\mathrm{MA}_{i,j} = \mathrm{InfGain}(i, j), \quad i \neq j. \tag{1}$$

For a committee of N beat trackers, they then calculate the $N(N-1)/2$ different mutual agreements and average them all to obtain the MMA. The researchers show how a low mean mutual agreement coincides with perceptual and musical properties that make tapping difficult for humans. They build a challenging dataset by selecting samples with MMA < 1 bit, given a committee of five beat trackers.

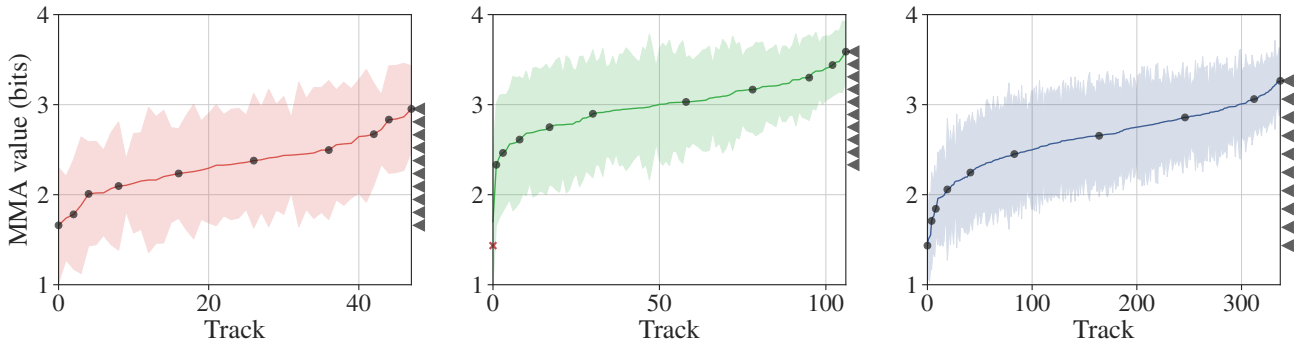
Later, in [42], they calculate the MaxMA, i.e., the algorithm whose output presents the maximum mutual agreement with the rest of the committee, showing that it provides the most reliable estimation for that given music example. They conduct subjective listening tests to determine a perceptual threshold for acceptable quality of this chosen output. This threshold is found to be 1.5 bits, for the same committee of five beat trackers. Their results also show a correlation between the test ratings and the MMA.

In this work, we followed a similar approach to select samples of various difficulties to the state-of-the-art algorithms. As mentioned in 4.1, we collected implementations of 14 beat tracking systems, removing some (the unavailable ones) featured in the original work [21], and adding others that were presented after its publication. We then extracted 30-second excerpts from all the different *sambas de enredo* in SAMBASET, and computed the MMA between estimations yielded by the beat trackers for all these 493 files. Figure 3 presents the ordered MMA values for excerpts from the three collections (HES, ESE, and SDE).

[6] Computed with standard settings using mir_eval [35] version 0.5.

Beat Tracker	CMLt (%)	AMLt (%)	F-meas. (%)	Inf. Gain (bits)
Aubio (AUB) [9]	59.4	65.6	61.9	2.30
BayesBeat-HMM (KR1) [25, 26]	42.7	65.6	67.6	2.27
BayesBeat-PF (KR2) [25, 27]	47.6	52.9	58.0	2.25
*BeatRoot (DIX) [17]	79.4	82.8	86.4	3.15
Davies (DAV) [15]	97.2	97.2	97.5	3.66
*Degara (DEG) [16]	88.3	91.2	89.7	3.40
*Ellis (ELL) [18]	76.9	76.9	78.7	3.35
IBT causal (IB1) [33]	83.4	83.4	86.2	2.45
*IBT non-causal (IB2) [33]	51.1	90.8	80.0	2.49
*Klapuri (KLA) [23]	61.3	63.7	63.1	3.09
BeatTracker (BO1) [5, 7]	98.1	98.1	98.6	3.78
DBNBeatTracker (BO2) [4, 26]	**99.5**	**99.5**	**99.5**	**3.80**
DBNDownBeatTracker (BO3) [6]	94.0	97.3	97.1	3.68
MultiFeature (MFT) [41]	86.4	86.4	86.8	3.55
Mean	76.1	82.2	82.2	3.09

Table 2: Ground truth performance of each beat tracking algorithm on the audio excerpts of SAMBASET. The best performance for each metric is highlighted in bold. The five-member committee proposed in [21] is indicated by an asterisk.



(a) HES collection (b) ESE collection (c) SDE collection

Figure 3: Collections sorted by MMA mean (solid line), with standard deviation (shaded region). Annotated samples (solid circles) were chosen as the closest to ten evenly spaced MMA values (solid triangles). One sample was treated as an outlier (cross) in (b).

For each collection, using the curve obtained from its excerpts, we determined P evenly spaced MMA values (including the maximum and minimum points), and selected the excerpts closest to each of these values to annotate. The reasons for this procedure are twofold: first, by selecting the same number of samples from each collection, we compensate for the large imbalance between them (e.g. in SDE there are nearly seven times more excerpts than in HES), while ensuring that their unique characteristics will be equally represented in the subset (recalling the variability in HES is much higher than that in ESE, or SDE); second, we guarantee that the algorithms will be compared within a group of samples to which they share different levels of consensus (and that would possibly provide a human annotator with a gamut of challenges). In total, thirty files were manually annotated ($P = 10$, i.e., ten from each collection), totalling just over 1 900 beats. It should be noted that a moderate number of annotated samples is sufficient, since we are dealing with a single music genre, which considerably limits the range of variations between them.

4.4 Discussion of Results

We see in Figure 3 that, in general, the fourteen beat tracking systems show more agreement in estimations for tracks in the ESE collection, followed by those in SDE, with HES in last. In fact, for over 50% of the tracks in ESE, the algorithms presented MMA > 3 bits, against slightly under 12% for SDE tracks and 0% in HES tracks in the same conditions. Considering an MMA > 2.5 bits, those percentages grow to 95%, 70% and 23%, respectively. This agrees with the authors' overall impression that the HES collection is the most "flavorful", whereas ESE is less expressive.

Regarding the test with the sampled subset, Table 2 gives the accuracy values for all algorithms, averaged over the thirty samples. Seven beat trackers perform better than the mean in all metrics, some of them outperforming the others by a large margin. In the end, the four best algorithms for our dataset are BO2, BO1, DAV, and BO3.

For the sake of comparison, we also evaluated the 493 excerpts with the five-member committee proposed in [21] and used in [42]: for 98.6% of the set the committee shows

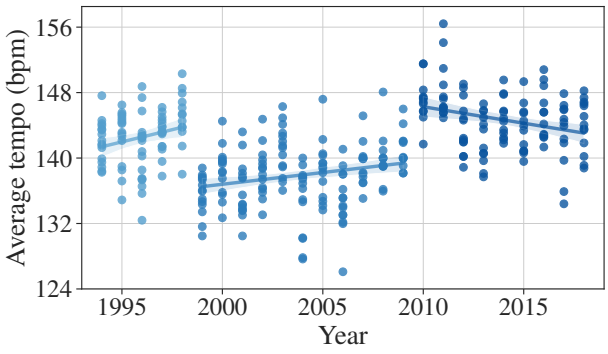


Figure 4: Variation of average tempo across the SDE collection with trend lines for three distinct regions and respective confidence intervals (shaded areas).

MMA > 1.5 bits; a single excerpt has MMA < 1 bit. This indicates that, overall, SAMBASET excerpts are not very challenging to the algorithms in this committee, which would provide a good number of acceptable estimations. Indeed, the good results shown by committee members in the ground truth performance suggests likewise. This analysis of state-of-the-art algorithms indicates a safe approach to semiautomatically annotating beats in this dataset.

5. MUSICOLOGICAL INSIGHTS

Here we investigate the evolution of average tempo in *samba de enredo* recordings across the years as represented in the SDE collection. For each excerpt, we compute the average tempo in beats per minute (bpm) as the inverse of the mean inter-beat interval, using the automatically detected beats. Figure 4 shows the average tempo for every track in SDE, plotted against the release year.

Although no clear trend is apparent from the whole data, we can readily verify the existence of local trends in three different regions of the graph. The first region accounts for the years of 1994 through 1998, and corresponds to the end of an era of "live" recordings in the *Teatro de Lona* (Barra da Tijuca), a large circus-like tent. As Moehn reveals on his essay "The Disc is not the Avenue" [31], by then the recordings were being made with a large number of musicians from each *escola* (around sixty) as well as large choirs from the respective community.

A radical change took place in the production of the 1999 disc: the entire process was moved to the studio and the number of *escola* members was reduced, not only to cut costs, but also to regain control over the sound organization [31]. Producers wanted the disc to sound "clear" and, thus, constrained the creative liberties of the *bateria*'s directors (e.g. they were not allowed to choose the tempo of the performance or to follow certain musical conventions that are common in a live performance). This was an attempt to recover the disc's marketability (sales had been dropping in previous years), despite distancing it from the actual phenomenon of the *samba de enredo* [31]. In 2010, "live" recordings were resumed, this time in the *Cidade do Samba* (Gambôa). Producers retreated in their interference on the soundscape creation, and the *escolas* were able to reclaim the final saying in some aspects of the recording, such as the tempo. The larger space provided by the *Cidade do Samba* also lead to an increase in the number of musicians taking part in the recordings: more than 8 000 for the 2014 CD against 1 500 in the 1998 recording [31].

Therefore, the first and third regions of Figure 4 more closely represent actual *samba de enredo* performances. In particular, notice that the average bpm in the third region is above the averages in the other two regions. This can be seen as a direct translation to the digital media of the decisions to accelerate the live performances (and the marching pace), so that the *escolas* satisfy changes in parading time limits, as reported by many specialists [12, 22, 34].

6. CHALLENGES

It would be very interesting to enrich this dataset with other types of annotations. In particular, one could think of generating ground truths for section boundaries (e.g. verses and the two different choruses that are very common in *samba de enredo* compositions), chord annotations (for instruments such as the *cavaquinho*), and instrument activity. As in the case of the CompMusic project [38], pitch contour annotations for soloist voices could be produced, as well as time-aligned lyrics and percussion transcription.

With these annotations, SAMBASET can provide many challenges to state-of-the-art algorithms in different MIR tasks. Tracks (specially in SDE) contain a plethora of simultaneous sounds of different qualities and textures, e.g. harmonic and percussive instruments, soloists and choirs. These could pose hard problems to vocal F0 or chord estimation systems. Also, singing voice annotation and lyrics could allow the study of soloist's interpretation as to phrasing, preferred ornaments, or characteristic syncopation; along with the metadata provided, it would be possible for example, to work on singer classification.

7. CONCLUSION

In this paper, we presented SAMBASET, a large *samba de enredo* dataset with rich metadata, beat and downbeat annotations. We provided a detailed overview of its contents [7] and reported a study on the performance of a group of state-of-the-art beat trackers over the set. We also motivated one musicological use of the dataset, i.e., the study of changes in *samba de enredo*'s rhythmic properties across several years. We expect that SAMBASET allows for technical improvements in traditional MIR tasks via new perspectives on problem solving that arise from contemplating cultures different from those to which we are accustomed.

8. ACKNOWLEDGEMENTS

Authors would like to thank CNPq and CAPES-ANII-CNRS (STIC AmSud program) for funding this work.

[7] Visit `http://www.smt.ufrj.br/~starel/sambaset/` for more information.

9. REFERENCES

[1] S. M. Araujo Junior. *Acoustic labor in the timing of everyday life: a critical contribution to the history of samba in Rio de Janeiro*. Phd thesis, University of Illinois, 1992.

[2] S. Arom. Structuration du temps dans les musiques d'afrique centrale: périodicité, mètre, rythmique et polyrythmie. *Revue de Musicologie*, 70(1):5–36, January 1984.

[3] S. Böck, F. Korzeniowski, J. Schlüter, F. Krebs, and G. Widmer. madmom: a new python audio and music signal processing library. In *24th ACM International Conference on Multimedia*, pages 1174–1178, Amsterdam, The Netherlands, October 2016.

[4] S. Böck, F. Krebs, and G. Widmer. A multi-model approach to beat tracking considering heterogeneous music styles. In *15th Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 603–608, Taipei, Taiwan, October 2014.

[5] S. Böck, F. Krebs, and G. Widmer. Accurate tempo estimation based on recurrent neural networks and resonating comb filters. In *16th Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 625–631, Málaga, Spain, October 2015.

[6] 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)*, pages 255–261, New York, USA, August 2016.

[7] S. Böck and M. Schedl. Enhanced beat tracking with context-aware neural networks. In *14th Int. Conf. on Digital Audio Effects (DAFx)*, pages 135–139, Paris, France, September 2011.

[8] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. R. Zapata, and X. Serra. ESSENTIA: an audio analysis library for music information retrieval. In *14th Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 493–498, Curitiba, Brazil, November 2013.

[9] P. M. Brossier. *Automatic Annotation of Musical Audio for Interactive Applications*. PhD thesis, Department of Electronic Engineering, Queen Mary University of London, 2006.

[10] C. Cannam, M. O. Jewell, C. Rhodes, M. Sandler, , and M. d'Inverno. Linked data and you: Bringing music research software into the semantic web. *Journal of New Music Research*, 39(4):313–325, 2010.

[11] C. Cannam, C. Landone, and M. Sandler. Sonic Visualiser: An open source application for viewing, analysing, and annotating music audio files. In *ACM Multimedia 2010 Int. Conf.*, pages 1467–1468, Florence, Italy, October 2010.

[12] F. L. Cunha. Samba locations: An analysis on the carioca samba, identities, and intangible heritage (Rio de Janeiro, Brazil). In F. L. Cunha, M. Santos, and J. Rabassa, editors, *Latin American Heritage*, The Latin America Studies Book Series, chapter 1, pages 3–20. Springer International Publishing, 2018.

[13] M. E. P. Davies, N. Degara, and M. D. Plumbley. Evaluation metrics for musical audio beat tracking algorithms. Technical Report C4DM-TR-09-06, Centre for Digital Music, Queen Mary University of London, London, UK, 2009.

[14] M. E. P. Davies, N. Degara, and M. D. Plumbley. Measuring the performance of beat tracking algorithms using a beat error histogram. *IEEE Signal Processing Letters*, 18(3):157–160, March 2011.

[15] M. E. P. Davies and M. D. Plumbley. Context-dependent beat tracking of musical audio. *IEEE Transactions on Audio, Speech and Language Processing*, 15(3):1000–1009, February 2007.

[16] N. Degara, E. A. Rúa, 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, January 2012.

[17] S. Dixon. Evaluation of the audio beat tracking system BeatRoot. *Journal of New Music Research*, 36(1):39–50, 2007.

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

[19] S. W. Hainsworth and M. D. Macleod. Particle filtering applied to musical tempo tracking. *EURASIP Journal on Advances in Signal Processing*, 15:2385–2395, November 2004.

[20] M. A. Hertzman. A Brazilian counterweight: Music, intellectual property and the African diaspora in Rio de Janeiro (1910s–1930s). *Journal of Latin American Studies*, 41(4):695–722, November 2009.

[21] 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, November 2012.

[22] IPHAN. *Matrizes do Samba no Rio de Janeiro: partido-alto, samba de terreiro, samba-enredo*. National Institute of Historic and Artistic Heritage (IPHAN), Brasília, Brazil, 2014. IPHAN dossier 10.

[23] 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, January 2006.

[24] M. Kolinski. A cross-cultural approach to metro-rhythmic patterns. *Ethnomusicology*, 17(3):494–506, September 1973.

[25] F. Krebs, S. Böck, and G. Widmer. Rhythmic pattern modelling for beat and downbeat tracking from musical audio. In *14th Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 227–232, Curitiba, Brazil, November 2013.

[26] F. Krebs, S. Böck, and G. Widmer. An efficient state-space model for joint tempo and meter tracking. In *16th Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 72–78, Málaga, Spain, October 2015.

[27] F. Krebs, A. Holzapfel, A. T. Cemgil, and G. Widmer. Inferring metrical structure in music using particle filters. *IEEE Transactions on Audio, Speech and Language Processing*, 23(5):817–827, May 2015.

[28] L. S. Maia, P. D. Tomaz Jr., M. Fuentes, M. Rocamora, L. W. P. Biscainho, M. V. M. Costa, and S. Cohen. A novel dataset of Brazilian rhythmic instruments and some experiments in computational rhythm analysis. In *2018 AES Latin American Congress of Audio Engineering*, pages 53–60, Montevideo, Uruguay, September 2018.

[29] B. McFee, E. J. Humphrey, and J. P. Bello. A software framework for musical data augmentation. In *16th Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 248–254, Málaga, Spain, October 2015.

[30] Brian McFee, C. Raffel, D. Liang, D. P. W. Ellis, M. McVicar, E. Battenbergk, and O. Nieto. librosa: Audio and music signal analysis in Python. In *14th Python in Science Conf. (SciPy)*, pages 18–24, Austin, USA, July 2015.

[31] F. J. Moehn. 'The disc is not the avenue': Schismogenetic mimesis in samba recording. In P. D. Green and T. Porcello, editors, *Wired for Sound: Engineering and Technologies in Sonic Cultures*, chapter 3, pages 47–83. Wesleyan University Press, 2005.

[32] L. Nunes, M. Rocamora, L. Jure, and L. W. P. Biscainho. Beat and downbeat tracking based on rhythmic patterns applied to the uruguayan candombe drumming. In *16th Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 264–270, Málaga, Spain, October 2015.

[33] J. L. Oliveira, F. Gouyon, L. G. Martins, and L. P. Reis. IBT: A real-time tempo and beat tracking system. In *11th Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 291–296, Utrecht, The Netherlands, August 2010.

[34] Y. Prado. Padrões musicais do samba-enredo na era do Sambódromo. *Música em Perspectiva*, 8(1):155–195, June 2015.

[35] C. Raffel, B. McFee, E. J. Humphrey, J. Salamon, O. Nieto, D. Liang, and D. P. W. Ellis. mir_eval: A transparent implementation of common MIR metrics. In *15th Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 367–372, Taipei, Taiwan, October 2014.

[36] D. M. Randel. *The Harvard Dictionary of Music*. Belknap Press of Harvard University Press, Cambridge, USA, 4 edition, 2003.

[37] C. Sandroni. *Feitiço Decente: Transformações do samba no Rio de Janeiro (1917-1933)*. Jorge Zahar Ed./Ed. UFRJ, Rio de Janeiro, Brazil, 2 edition, 2008.

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

[39] C. N. Silla Jr., A. L. Koerich, and C. A. A. Kaestner. The Latin music database. In *9th Int. Conf. on Music Information Retrieval (ISMIR)*, pages 451–456, Philadelphia, USA, September 2008.

[40] J. M. Sousa, E. T. Pereira, and L. R. Veloso. A robust music genre classification approach for global and regional music datasets evaluation. In *2016 IEEE Int. Conf. on Digital Signal Processing*, pages 109–113, Beijing, China, October 2016.

[41] J. R. Zapata, M. E. P. Davies, and E. Gómez. Multi-feature beat tracking. *IEEE Transactions on Audio, Speech and Language Processing*, 22(4):816–825, April 2014.

[42] J. R. Zapata, A. Holzapfel, M. E. P. Davies, J. L. Oliveira, and F. Gouyon. Assigning a confidence threshold on automatic beat annotation in large datasets. In *13th Int. Soc. for Music Information Retrieval Conf. (ISMIR)*, pages 157–162, Porto, Portugal, October 2012.