INVESTIGATING TIME-LINE-BASED MUSIC TRADITIONS WITH FIELD RECORDINGS: A CASE STUDY OF CANDOMBLÉ BELL PATTERNS

Lucas S. Maia[1,2] **Richa Namballa**[3] **Martín Rocamora**[4,5]
Magdalena Fuentes[3] **Carlos Guedes**[1]
[1] MaSC, New York University Abu Dhabi, UAE
[2] PEE/COPPE, Universidade Federal do Rio de Janeiro, Brazil
[3] New York University, United States
[4] Universitat Pompeu Fabra, Spain [5] Universidad de la República, Uruguay
{lucas.maia, richa.namballa}@nyu.edu

ABSTRACT

We introduce a series of transdisciplinary corpus studies aimed at investigating cross-cultural trends in time-line-based music traditions. Our analyses concentrate on a compilation of field recordings from the Centre de Recherche en Ethnomusicologie (CREM) sound archive. To demonstrate the value of an interdisciplinary approach combining ethnomusicology and music information research to rhythmic analysis, we propose a case study on the bell patterns used in the musical practices of Candomblé, an Afro-Brazilian religion. After removing vocals from the recordings with a deep learning source separation technique, we further process the instrumental segments using non-negative matrix factorization and select the bell components. Then, we compute a tempo-agnostic rhythmic feature from the bell track and use it to cluster the data. Finally, we use synthesized patterns from the musicological literature about Candomblé as references to propagate labels to the rhythmic clusters in our data. This semi-supervised approach to pattern analysis precludes the need for downbeat and cycle annotations, making it particularly suited for extensive archive investigations. Lastly, by comparing bell patterns in Candomblé and a West African music tradition, we lay the foundation for our future cross-cultural research and observe the potential application of this methodology to other time-line-based music.

1. INTRODUCTION

Over the years, the music information retrieval/research (MIR) community has embraced more culturally-inclusive research geared towards the analysis of non-Western music [1]. Many of these studies have done extensive work to increase the representation of certain musical styles from specific cultures [2]. However, few MIR initiatives have attempted to investigate multiple musical cultures simultaneously at a larger scale due to the level of difficulty and lack of data. These hardships reinforce a cycle of under-representation of various populations and risk further emphasizing the perspective of MIR through a Western-centric lens.

In this paper, we propose one of several studies aimed at a more global, inclusive, and transdisciplinary approach to MIR that puts humanistic (ethnomusicological and anthropological) and computational approaches in dialogue within a framework defined as Sonic Digital Humanities [3]. Our work centers on a substantial corpus of data from the Centre de Recherche en Ethnomusicologie (CREM). While we will explore only a part of this archive in this paper, we discuss its contents and their importance in Section 3. In Sections 4-7, we introduce a preliminary investigation on a subset of the archive analyzing bell patterns in Brazilian Candomblé and music from West Africa to demonstrate the potential this data has for future cross-cultural research on time-line-based acoustic traditions.

2. RELATED WORK

Ethnomusicology and MIR are often associated in a way that one is viewed as the source discipline while the other is the target [4]. However, researchers have suggested treating them as partners rather than as a hierarchy [5, 6]. With this approach, the MIR community can develop new computational methodologies which incorporate external information, such as cultural context, to better understand audio signals [1]. Specifically, we apply a framework known as the Sonic Digital Humanities (SDH) to our investigation. SDH is a branch of the Digital Humanities concerned with digital collections of music and other forms of sonic culture. It provides a space in which computational means to the analysis of sound culture may be developed and carried out in a productive dialogue with humanistic modes of data collection and critical inquiry [3].

By analyzing collections from an SDH perspective, we intend to ask questions about the cross-cultural relationships of different musical styles on a global scale. While engaging in data-driven analyses of these musical styles, we attempt to understand (1) whether these findings support evidence collected from ethnomusicological studies about cross-cultural influences and (2) how these large-scale computational investigations can provide further insights into comparing music cross-culturally. In the case of (1), validating (or not) musicological evidence about cross-cultural relationships can improve MIR approaches to provide more reliable large-scale studies of lesser-known styles in the digital world. With (2), these methods can yield new findings of what characterizes a musical style born out of cross-cultural influence.

The first steps in expanding MIR beyond the exclusive sphere of Western music involved improving cultural representation in datasets. As a result, several culture- or style-specific corpora have been published for studies on music such as American ragtime, Beijing opera, and more [7, 8]. These advancements have improved access to a variety of data, but have yet to make a dent in the dominance of Western methodologies in MIR. A recent push by Huang et al. [9] for the MIR community to go beyond the collection of diverse data, collaborate with musicologists, and reflect on the way we engage with music serves as an appropriate objective for our studies of the CREM archive.

Despite the variety and richness of the information available in the CREM archive, very few studies have been published about this data. One MIR study utilizes the database to evaluate a proposed timbre classification method on a diverse set of musical instruments with the intent of allowing the indexing of ethnomusicological databases [10]. This sparseness of research presents major opportunities to explore the CREM archive in greater detail over a series of long-term, novel studies.

Our first investigation concerns the analysis of time lines, also known as bell patterns, in large collections of music. Time lines are short, cyclic patterns played in ostinato, often with a bell, castanet, or sticks [11, 12], that are used as a "controlling structural concept" [13, p. 1] in African music. This type of organization extends beyond geographical boundaries and can be heard in Afro-diasporic musical styles from the Caribbean or South America, for example. A key aspect of time lines is that they are qualitatively different from the concept of meter, as they originate from the movement of feet in dance [11], and denote a circularity that is characteristic of African music traditions [13]. This distinguishing factor calls for computational pattern recognition strategies that go beyond traditional methods of meter detection.

Toussaint proposed several mathematical methods, including geometric and graphical ones, for the analysis of clave-bell rhythm time lines [14]. Despite not being originally automated, they served as foundational work for future research concerning rhythmic complexity. Soon after, Toussaint continued their work on clave-bell time lines by comparing metrics for rhythmic similarity, such as Hamming distance and Euclidean interval vector distance [15]. All of the methods described in [14] and [15] require manual annotations, which are time-consuming and not scalable to a corpus as vast as the CREM archive.

Consequently, our proposed pipeline for time line pattern analysis is semi-supervised, with the majority of feature extraction and similarity computations automated. We draw inspiration from [16], who used template matching to track tempi of Afro-Cuban clave rhythms. However, their method has the drawback of requiring an exhaustive search for matching every tempo at each onset. Additionally, we consider the approach by [17], who inferred meter from Candombe recordings using rhythmic templates learned with the help of annotations. We improve upon these methodologies by using reference tracks to compute similarity measures based on [18]. The scale transform magnitudes (STM) [18] operate on the autocorrelation of the signal's onset strength. They are robust to tempo variation, which facilitates the transfer of labels from the references to the tracks under study.

In this paper, we use Candomblé as a case study of time-line-based music. Candomblé is an Afro-Brazilian religion known for syncretically combining elements from many cultures, most notably Yoruba, Bantu, and Fon — which were brought to Brazil by enslaved West African populations [19]. Music plays a crucial role in the religious practices of Candomblé. Antiphonal songs are performed throughout the entire ceremony, accompanied by a drumming ensemble, always with the intent of allowing the participants and certain deities (*orixás*) to communicate [20]. Different rhythmic patterns, in both singing and drumming, are associated with different *orixás*. There are a few historical collections of Candomblé field recordings in the CREM archive, serving as a valuable resource of audio data for our investigation.

3. CREM-NYUAD COLLECTION

The CREM database is an extensive archive of digitized audio recordings from cultures around the world. Spanning from the beginning of the 20th century to today, the archive contains over 48,000 field recordings and more than 17,000 published commercial recordings representing over 1300 ethnic groups across 199 countries. The public has access to rich metadata cataloguing the database as well as thousands of recordings available to listen to for free on the archive's website. [1]

Through a partnership with the Centre National de la Recherche Scientifique (CNRS), New York University Abu Dhabi (NYUAD) has acquired a subset of the CREM archive for the purpose of analyzing the sound recordings. Henceforth, we call this subset the CREM-NYUAD collection. The CREM-NYUAD collection consists of 14,379 records from 129 countries, with a majority coming from Africa, Asia, and South America. In particular, Vietnam, Nepal, Madagascar, Gabon, and Algeria are among the countries with the most records in the dataset. Each item consists of audio features, such as spectrograms and tempograms, extracted during a prior collaboration between CREM and NYU. The associated metadata for each record contains basic data about the item, such as the collec-

[1] https://archives.crem-cnrs.fr

tion name and date, but also includes valuable information about the location, language, instrumentation, and ethnographic context of the recording. This information provides a significant advantage in our pilot study of Candomblé bell patterns, as we will see in Sections 4-7.

An important feature of our collection, which distinguishes it from many other datasets used in MIR research, is the prevalence of field recordings in the corpus. Field recordings provide important cultural context through the settings in which the traditions are recorded, such as in the *terreiros* (places of worship) of Candomblé [21]. In contrast to commercial studio tracks, field recordings are often taken in natural conditions where there are various social and environmental sounds, as well as noise, in the final recording [22, 23]. Furthermore, the time span over which these field recordings are collected often reflects the technological progress of the time period with more recent recordings producing higher quality audio. The acoustic diversity of the collection poses additional challenges to our computational methods in the form of silence, noise, and artifacts (e.g., clicks). We attempt to overcome some of these obstacles to the analysis in our pipeline, but save any audio restoration endeavors for future work.

4. BELL PATTERNS IN CANDOMBLÉ

The drumming ensemble in Candomblé typically consists of three differently-sized drums called *atabaques*, a dried gourd covered in beads known as *xequerê*, and a single or double clapperless bell called *gan* or *agogô*. Figure 1 shows ten essential Candomblé bell patterns as notated in [24]. These motifs were identified as the main patterns utilized by the Ketu nation, the largest branch of the Candomblé religion today. For example, pattern 1 represents the bell part in *vassi*, which is a common pattern in many different rhythms of Candomblé and is performed by bells with two accompanying drums. Pattern 2 is the same as the Son clave [14]. Pattern 3, known as *ijexá*, gained popularity in not only religious contexts, but also in the festive *Carnaval* parades held in Salvador, Bahia.

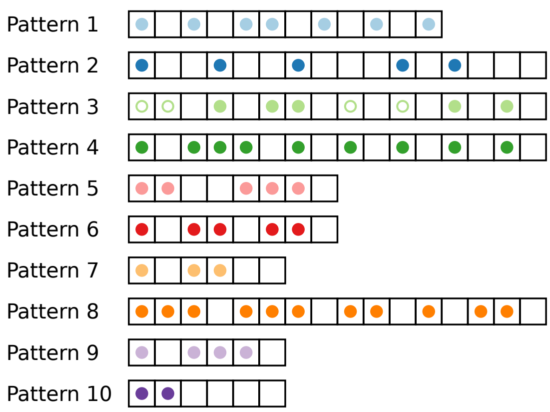

Figure 1: Examples of Candomblé bell patterns in time unit box system (TUBS) notation [24]. Open and closed dots indicate the use of high- and low-pitched bells.

5. METHODOLOGY

The process we employ to extract rhythmic features associated with bell patterns in field recordings is encapsulated in Figure 2, and explained further in this section.

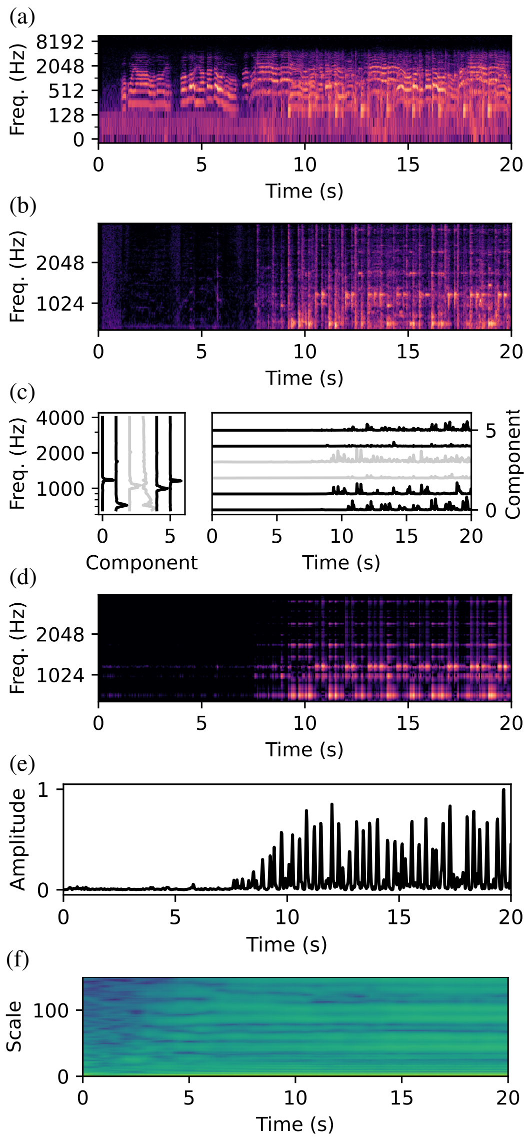

Figure 2: Workflow for extracting rhythmic features for bell patterns: (a) spectrogram of the original track (vocals and instruments); (b) the isolated instrumental part, from 650 to 4000 Hz; (c) NMF-learned templates and activations (highlighting those selected via the spectral crest); (d) reconstructed spectrogram of the bell part; (e) smooth bell activation function; and (f) frame-wise STM.

5.1 Source Separation

In order to analyze the bell patterns in greater detail, we need to isolate them from the remainder of the audio track. To do so, we first remove the vocals from each track due to their potential obstruction of the bells' frequency band. Using the pre-trained state-of-the-art hybrid transformer source separation model, Demucs [25], we preserve the instrumentals by separating the vocal stem.

To further abstract the bell patterns from the rest of the rhythm sections, we decompose the non-vocal track with non-negative matrix factorization (NMF) [26]. First, we resample the non-vocal signal to 8000 Hz. Then, we compute its short-time Fourier transform (STFT) with a 64-ms window and a hop size of 20 ms. We restrict the NMF analysis from 650 to 4000 Hz, which discards the main frequencies from low-pitched drums. At this time, we assume that the signal contains primarily the bell tones (sometimes in two distinct pitches, when both bells of the *agogô* are used) and noise-like components emanating from the *xequerê* or the other drums' attacks. For this reason, we run the NMF algorithm with $n = 6$ components. After the algorithm converges, we use the spectral crest [27] to identify sources corresponding to the bells by selecting components whose templates are more tonal in nature. The geometric mean of all crest factors serves as a threshold. Finally, we reconstruct the separated spectrogram of the bell part with the dot product of the matrices composed by the selected template–activation pairs.

5.2 Feature Extraction

The next step in our workflow is computing a rhythmic feature based on the scale transform magnitudes [18]. We first compute the time derivative of the log-compressed reconstructed source spectrogram, using a factor of compression $C = 1000$ as in [28]. All of the bins are summed up, and we apply half-wave rectification to keep only positive peaks. To smooth this accent signal, we use a lag of 3 frames in the computation of the time-difference [29] and further process the signal by convolving it with a Gaussian kernel ($\sigma = 20$ ms). Finally, we follow the procedure of [18] and determine the local autocorrelation of the accent signal with an 8-second moving rectangular window (hop size of 0.5 s). The direct scale transform [30] converts the autocorrelation at each frame into the scale domain, such that tempo is not encoded in the representation, and we keep only the first 150 scale coefficients. We discard all frames at the start of the signal whose energy lies below a threshold of -60 dB. Feature vectors can be compared using cosine similarity or Euclidean distance [18], with the former being better suited for handling changes in level between the recordings.

5.3 Label Propagation

We follow a semi-supervised procedure to classify patterns in the dataset. For this purpose, we create synthetic versions of the patterns in Figure 1 with no accents or timing deviations. We extract the rhythmic patterns of these synthesized reference tracks using the same pipeline as before. The only differences are that we use all $n = 3$ NMF components to generate a single activation and that we summarize the STM feature by taking the average along the time axis. Lastly, we propagate labels to the original (unlabeled) dataset in the following fashion:

1. For each track i in the dataset, we measure the maximum pairwise distance, σ_i, between STM frames;
2. For each frame j of track i, we find the closest data point, r_k, from the reference set, such that the distance $d(x_{ij}, r_k)$ is minimal;
3. If $d(x_{ij}, r_k) \leq \sigma_i$, x_{ij} receives the same label as r_k, else it receives a "null" label;
4. "Winner-take-all": we perform plurality voting among all labels for x_i where the most prevalent label is used to represent the entire track.

While the labels can be propagated within the feature space, this process can also be intuitively performed in a lower-dimension embedding space.

6. EXPERIMENTS AND RESULTS

We select a specific set of tracks from the CREM-NYUAD collection on which to run our entire pipeline. We rely on the metadata described in Section 3 to identify which files contain bell sounds. Table 1 shows the countries in West Africa (and Brazil) with these bell patterns and the number of recordings from each country.

Country	# Records
Congo-Brazzaville	98
Brazil	71
Benin	42
Angola	30
Gabon	27
Mali	24
Côte d'Ivoire	11

Table 1: Number of records with bell patterns per country.

In this initial experiment, we further restrict our scope by selecting, from those files recorded in Brazil, a set of recordings by ethnographer (and Candomblé initiate) Pierre Verger. [2] Moreover, with the assumption that bells mostly establish a cyclic pattern, we consider only the first 60 s of each recording. Next, we proceed with the analysis from Section 5; i.e., after our pre-processing steps, we extract the rhythmic features for all tracks in the subset and in the reference set. We then perform label propagation from the reference set to the subset. Using UMAP [31], a manifold learning technique, we present the results in Figure 3.

With regards to structure, our pipeline clearly extracts meaningful information from the rhythmic patterns in the subset, as many distinct clusters are visible. Interestingly, we observe that the reference patterns are well distributed among these clusters.

The label propagation procedure also reveals important aspects of the data distribution. For instance, we notice that the approach labels a large portion of the frames as belonging to the "pattern 1" archetype. Furthermore, note how closely patterns 7 and 9 are represented in the embedding. This proximity is easily explained by them differing on only a single beat (see again Figure 1). The families of patterns 2 and 5 also appear near each other in the manifold, but this time their pattern lengths are unequal. However, by "interpolating" pattern 5 and cyclically rotating

[2] `https://archives.crem-cnrs.fr/archives/collections/CNRSMH_I_2007_011/`

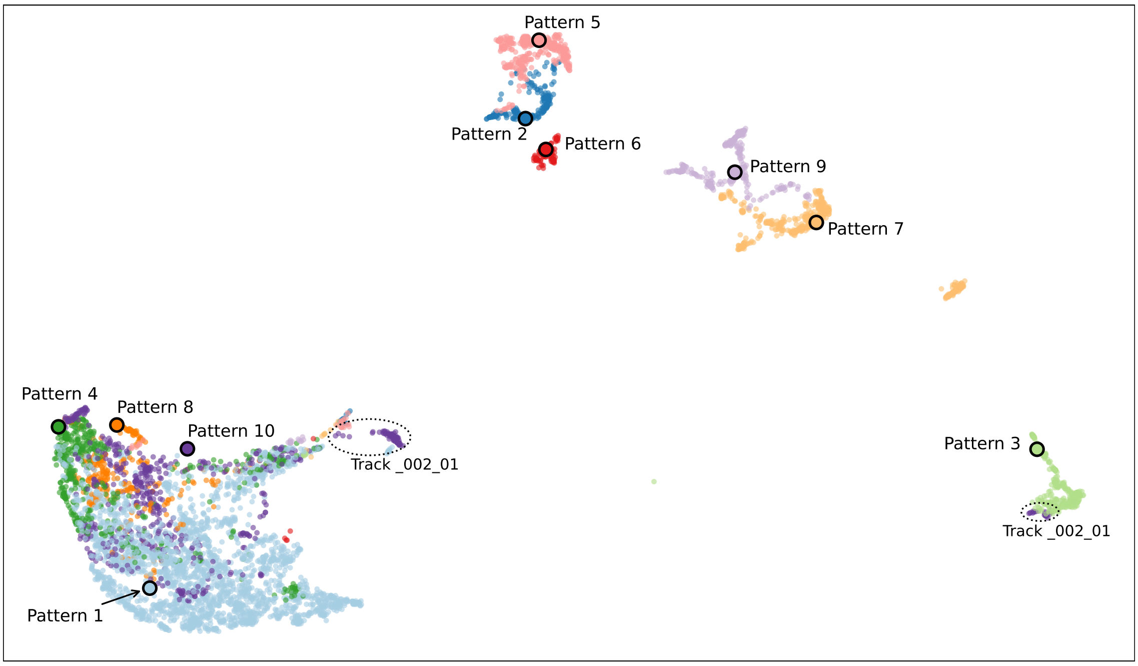

Figure 3: UMAP projection of the Candomblé bell patterns and reference patterns (cosine metric, n-neighbors = 70, min-dist = 0.1). The same color coding of Figure 1 is used here. Circled: the frames corresponding to track `_002_01`.

pattern 2 (Figure 4), we see that they are more alike than they appear at the surface level, especially considering that the STM feature is independent of tempo. A similar argument explains the close proximity of clusters for patterns 5 and 6.

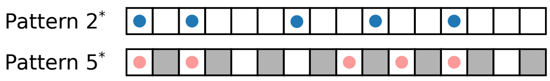

Figure 4: Interpolation and cyclical rotation of patterns 5 and 2 reinforce the similarities found in the embedding. Three of the five onsets match, and the remaining two are a slight shift away from their counterparts.

We can also assess the label propagation method's performance by investigating some information from the metadata. For example, recordings `_002_01`, `_004_03`, `_005_04`, and `_005_05` contain the indication "*ijexá* rhythm" ("*rhythme ijexá*") in their titles, so presumably they correspond to the same type of bell pattern as pattern 3. Consequently, in the embedding, their frames are clustered and labeled together as part of the "pattern 3" archetype. The only exception is recording `_002_01` (circled), which was incorrectly classified as pattern 10 and divided into two sections: the majority of frames are situated in the easternmost region of the large 1–4–8–10 cluster, while a handful of remaining frames are found near pattern 3. In this case, the misclassification could be attributed to the crest selection procedure's inability to retrieve the main component of the highest-pitched bell.

Figure 5 showcases selected examples of the onset activations from the subset, juxtaposed with their corresponding reference activations. To ensure alignment, we manually adjusted the references' timing to match the excerpts. Differences between the references and audio realizations, such as additional or missing notes, can be ascribed to recording conditions or introduced by the pipeline. Despite these discrepancies, the workflow demonstrates robustness, as evidenced by the confirmation of the automatic classification through listening tests. Pattern variations can also originate from the player, who may miss a note or add embellishments (flams). Another type of discrepancy we have identified, resulting from small scale deviations, is illustrated with recording `_004_09`.

Lastly, we conduct another analysis which uses a larger number of recordings with bell patterns from the Republic of the Congo (see Table 1). We follow the same procedure as before, but lower the minimum frequency for the NMF decomposition from 650 to 300 Hz, since West African bells are typically larger and lower pitched. Figure 6 displays the embedding of both the Brazilian and Congolese patterns from our subset. This visualization shows similarities between some of the patterns; these potential cross-cultural intersections require further investigation. In particular, with a "plurality voting" procedure similar to our label propagation scheme, we can detect that recording `_030_03` from collection `CNRSMH_I_1974_013`[3] is the most akin to the patterns of the Brazilian recordings. A short listening test confirms that the bell in this recording performs a rhythmic pattern similar to that of pattern 1 (the most common one in the Brazil subset).

7. DISCUSSION

We emphasize two important consequences of our study. Both observations emerge from our attempt to balance

[3] `https://archives.crem-cnrs.fr/archives/items/CNRSMH_I_1974_013_030_03/`

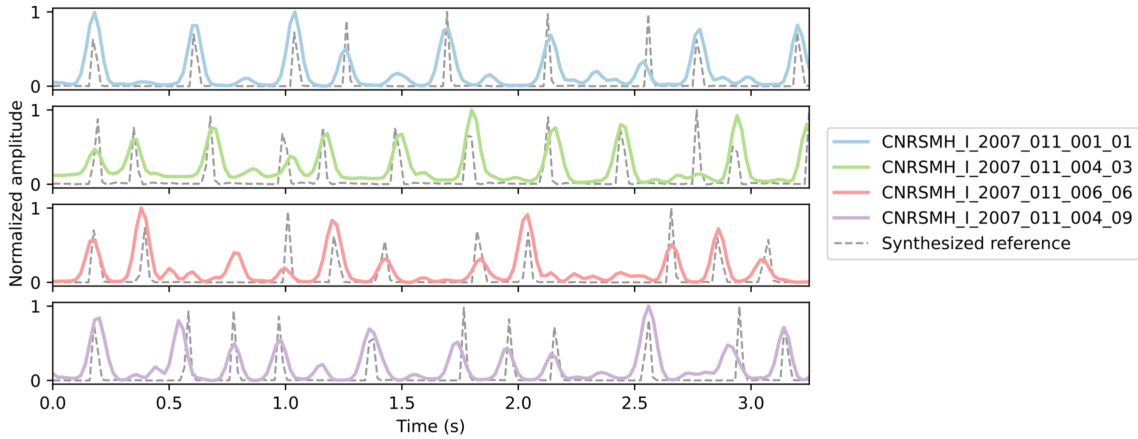

Figure 5: Excerpts of onset activations and their corresponding references. The same color coding of Figure 1 is used here.

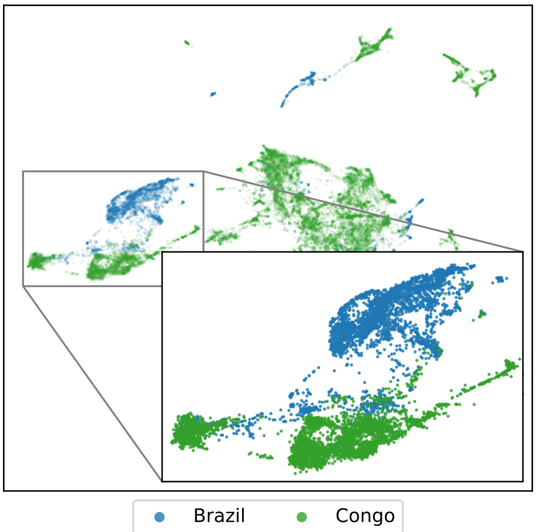

Figure 6: Embedding of rhythmic features from Brazilian and Congolese bell patterns in the subset.

data-driven methodologies with more humanistic perspectives using the SDH framework.

Firstly, the non-negative matrix factorization presents a significant bottleneck in our analysis. Ideally, we would use source separation to isolate the bells from the track directly. However, the out-of-the-box Demucs model is trained only to separate 4-6 stems of specific Western instruments. A blind application of this model often results in an unpredictable placement of the bell patterns, as they may end up in the drum stem for one track, but in the "other" category for another. This inconsistency underscores the need for a more culturally-inclusive, tailored approach, such as the few-shot source separation model proposed in [32], or even a new perspective on source separation as a task. Improvements such as these could offer a more accurate and consistent separation of the bells, enhancing the rhythmic salience.

Secondly, despite the technical challenges produced by field recordings, our feature extraction pipeline has proven to be remarkably robust. Most importantly, it respects the unique characteristics of West African and Afro-diasporic music, particularly the concept of rhythmic cycles. Time lines are qualitatively different from the concept of meter as a temporal hierarchical grouping mechanism and serve culture-specific purposes depending on the context of their use. While they can be mapped into meters due to their cyclic or recurrent nature, they often play "against" their metrical grid [12]. Anku [13] suggests that they should be perceived as a "circular concept" rather than a linear one, allowing performers to seamlessly enter and exit the performance with little inhibition. Our pipeline, which makes no assumptions regarding the notions of meter or downbeat and uses no annotations, was designed to respect these unique characteristics. The only attribute we infer is the cyclic nature of the rhythmic patterns to ease our computational load. Our minimal suppositions demonstrate the capacity of our methodology to expand to other styles of cyclic music, beyond what is studied in this paper.

8. CONCLUSION

We presented a pilot study investigating bell patterns in Candomblé from historical field recordings in a subset of the CREM archive, the CREM-NYUAD collection. Our approach is a preliminary venture in following the Sonic Digital Humanities (SDH) framework, to address the inherent challenges and complexities of applying computational methods to musical traditions which have been underrepresented in MIR. SDH aims to combine computational and ethnographic approaches in a dialogue on the same plane, while embracing any tensions which may entail in an agonistic fashion to push the traditional boundaries of interdisciplinarity [6].

Our study was influenced by the distinct characteristics of West African and Afro-diasporic music. Without requiring meter annotations, we could detect and classify patterns from the collection using a robust and adaptable pipeline, despite encountering challenging recording conditions and unique rhythmic structures. Our code is available at `https://github.com/nyuad-masc/crem-time-lines`. Future work will focus on addressing shortcomings and further refining our methods (e.g., source separation) to improve the analysis of time-line-based music traditions and allow the study of cross-cultural influences.

9. ETHICS STATEMENT

Our highest priority in this study is to respect the cultural heritage and to protect the privacy of the communities represented in the CREM-NYUAD collection. We recognize that every culture and tradition has nuances which cannot be captured by computational methods alone and should not be subjected to reductions or generalizations. Any analyses implemented using the materials from this corpus are solely for academic purposes in an effort to increase cultural diversity by establishing a dialogue between the humanities and music information research.

10. ACKNOWLEDGEMENTS

The authors would like to thank Mark Cartwright for his prior work extracting audio features from the CREM database, which serves as a crucial foundation for this and future research. We also thank Joséphine Simonnot, whose efforts were instrumental in the curation and acquisition of the CREM-NYUAD collection.

11. REFERENCES

[1] X. Serra, "A multicultural approach in music information research," in *Proceedings of the 12th International Society for Music Information Retrieval Conference*, Miami, USA, Oct. 2011, pp. 151–156.

[2] ——, "Creating research corpora for the computational study of music: the case of the CompMusic project," in *Proceedings of the AES 53rd International Conference on Semantic Audio*, London, UK, Jan. 2014, pp. 1–9.

[3] A. Eisenberg and C. Guedes, "Prolegomena for sonic digital humanities," in *British Forum for Ethnomusicology & International Council for Traditional Music Ireland Joint-Annual Conference*, Cork, Ireland, Apr. 2024.

[4] K. Neubarth, M. Bergeron, and D. Conklin, "Associations between musicology and music information retrieval," in *Proceedings of the 12th International Society for Music Information Retrieval Conference*, Miami, USA, Oct. 2011, pp. 429–434.

[5] P. Proutskova, "Musical memory of the world-data infrastructure in ethnomusicological archives," in *Proceedings of the 8th International Conference on Music Information Retrieval*, Vienna, Austria, Sep. 2007, pp. 161–162.

[6] G. Born, "Diversifying MIR: Knowledge and real-world challenges, and new interdisciplinary futures," *Transactions of the International Society for Music Information Retrieval*, vol. 3, no. 1, p. 193–204, 2020.

[7] A. Volk and W. B. de Haas, "A corpus-based study on ragtime syncopation," in *Proceedings of the 14th International Society for Music Information Retrieval Conference*, Curitiba, Brazil, Nov. 2013, pp. 163–168.

[8] R. Caro Repetto and X. Serra, "Creating a corpus of jingju (Beijing opera) music and possibilities for melodic analysis," in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, Taipei, Taiwan, Oct. 2014, pp. 313–318.

[9] R. Huang, A. Holzapfel, B. Sturm, and A.-K. Kaila, "Beyond diverse datasets: Responsible MIR, interdisciplinarity, and the fractured worlds of music," *Transactions of the International Society for Music Information Retrieval*, vol. 6, no. 1, pp. 43–59, 2023.

[10] D. Fourer, J.-L. Rouas, P. Hanna, and M. Robine, "Automatic timbre classification of ethnomusicological audio recordings," in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, Taipei, Taiwan, Oct. 2014, pp. 295–300.

[11] K. Agawu, *The African Imagination in Music*. Oxford University Press, 2016.

[12] ——, "The metrical underpinnings of African timeline patterns," 2013. [Online]. Available: https://www.youtube.com/watch?v=8ypTYNGLr5A

[13] W. Anku, "Circles and time: A theory of structural organization of rhythm in African music," *Music Theory Online*, vol. 6, no. 1, pp. 1–8, 2000.

[14] G. T. Toussaint, "A mathematical analysis of African, Brazilian, and Cuban clave rhythms," in *Bridges: Mathematical Connections in Art, Music, and Science*, 2002, pp. 157–168.

[15] ——, "A comparison of rhythmic similarity measures," in *Proceedings of the 5th International Conference on Music Information Retrieval*, Barcelona, Spain, Oct. 2004.

[16] M. Wright, W. A. Schloss, and G. Tzanetakis, "Analyzing Afro-Cuban rhythms using rotation-aware clave template matching with dynamic programming." in *Proceedings of the 9th International Society for Music Information Retrieval Conference*, Philadelphia, USA, Sep. 2008, pp. 647–652.

[17] 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 *Proceedings of the 16th International Society for Music Information Retrieval Conference*, Málaga, Spain, Oct. 2015, pp. 246–270.

[18] A. Holzapfel and Y. Stylianou, "Scale transform in rhythmic similarity of music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 1766–185, 2011.

[19] R. Bastide, *As Américas negras: as civilizações africanas no Novo Mundo*. São Paulo: Difel/Edusp, 1974.

[20] P. A. McGrath-Kerr, “A cannibalist’s manifesto: Candomblé rhythms for drum kit,” Ph.D. dissertation, Faculty of Fine Arts and Music, University of Melbourne, 2019.

[21] W. W. Megenney, “Afro-Brazilian percussion instruments: Etymologies & uses,” *Revista del CESLA. International Latin American Studies Review*, no. 9, pp. 25–35, 2006.

[22] M. Marolt, “Probabilistic segmentation and labeling of ethnomusicological field recordings,” in *Proceedings of the 10th International Society for Music Information Retrieval Conference*, Kobe, Japan, Oct. 2009, pp. 75–80.

[23] P. Proutskova and M. A. Casey, “You call that singing? ensemble classification for multi-cultural collections of music recordings.” in *Proceedings 10th of the International Society for Music Information Retrieval Conference*, Kobe, Japan, Oct. 2009, pp. 759–764.

[24] H. Schroy and B. Reis, “A orquestra do Candomblé da nação Ketu,” 2011. [Online]. Available: https://www.youtube.com/watch?v=kUgdwkBD-P8

[25] S. Rouard, F. Massa, and A. Défossez, “Hybrid transformers for music source separation,” in *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing*, Rhodes Island, Greece, Jun. 2023, pp. 1–5.

[26] D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Advances in Neural Information Processing Systems*, vol. 13, Denver, USA, 2000, pp. 556–562.

[27] G. Peeters, “A large set of audio features for sound description (similarity and classification) in the CUIDADO project,” IRCAM, Tech. Rep., 2004.

[28] P. Grosche and M. Muller, “Extracting predominant local pulse information from music recordings,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1688–1701, 2011.

[29] S. Böck and G. Widmer, “Maximum filter vibrato suppression for onset detection,” in *Proceedings of the 16th International Conference on Digital Audio Effects*, Maynooth, Ireland, 2013.

[30] W. J. Williams and E. J. Zalubas, “Helicopter transmission fault detection via time-frequency, scale and spectral methods,” *Mechanical Systems and Signal Processing*, vol. 14, no. 4, pp. 545–559, 2000.

[31] L. McInnes, J. Healy, N. Saul, and L. Großberger, “UMAP: Uniform manifold approximation and projection,” *The Journal of Open Source Software*, vol. 3, no. 29, p. 861, 2018.

[32] Y. Wang, D. Stoller, R. M. Bittner, and J. Pablo Bello, “Few-shot musical source separation,” in *Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing*, Singapore, Singapore, May 2022, pp. 121–125.