Internship Report: Rhythmic Pattern Analysis

Stefano Scola

July 2024

1 Introduction

As a research intern at the Music Technology Group, my primary objective was to explore and implement descriptors that capture rhythmic similarity between rhythmic patterns.

Over the past three months, I focused on the Scale Transform Magnitude (STM) [4] for rhythmic pattern recognition. My work covered two main aspects: demonstrating the usefulness of the feature by reproducing simple experiments previously presented in other publications, and implementing the feature in Python by leveraging existing libraries such as librosa.

2 Computation of STM

The Scale Transform Magnitude (STM) is a powerful feature used for analyzing rhythmic structure and measuring rhythmic similarity. Its key advantages lie in its invariance to tempo changes and its ability to function without beat annotations.

The implementation of STM involves the computation of the following:

1. Onset Strength Signal
2. Short-Time Auto-Correlation
3. Scale Transform

A. Scale Invariant Rhythm Descriptor

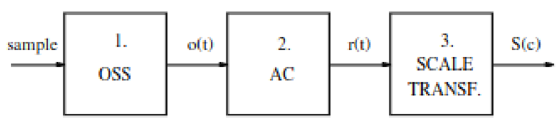

Fig. 1. Computational steps of scale invariant rhythm descriptors

Figure 1: Steps needed for the computation of the scale transform magnitude

Let's dive into the details of the feature computation:

1. The input audio is downsampled to 8kHz to improve efficiency and because high resolution is not necessary for rhythm analysis.

2. The spectrogram used to calculate the onset strength signal is computed with a Hann window of length 256 and a hop size of 160. It is then mapped to 40 mel bands, and finally, the amplitude is converted to dB.

3. The short-time auto-correlation is computed using a window (Hamming or rectangular) with specific time-lag parameter. Following the methodology presented in [6], a Hamming window of size 8 seconds with a 0.5-second hop size has been employed.

4. To introduce scale-invariant properties, the auto-correlation is transformed using either the Mellin Transform [1] or the Direct Scale Transform [10].

5. Finally, to obtain a 1-dimensional representation, the feature is averaged across all frames, and the coefficients at the end are discarded.

3 Demo Notebooks

During the project, I implemented several notebooks to highlight the computational steps, assess the scale invariance property of the feature, and showcase its application with machine learning techniques.

3.1 Classification and Clustering

The majority of my time was spent replicating previous experiments that employ the STM as a feature for classification or clustering tasks.

For classification, I performed genre classification on two datasets: BallRoom [3] and Cretan dances, as originally presented in the paper where the feature was hypothesized [4]. However, it's important to note that using music genre classification as a proxy task to validate and investigate rhythm can be uninformative and sub-optimal [2]. I used a simple k-Nearest Neighbors (KNN) algorithm to match the methodology presented in the original paper.

Given the limited reliability of music genre labels in terms of rhythmic patterns, I proceeded to conduct a clustering experiment using unlabeled data. It's worth noting that the data employed comes from different continental music traditions.

For clustering, I employed K-means and K-medoids algorithms, along with UMAP and t-SNE for dimensionality reduction and visualization, following the methodology presented in [6]. The data used consisted of:

1. The Groove MIDI dataset, containing beats along with their respective style labels.

2. Unlabeled data representing musical traditions from distinct parts of the world, including African and South American music traditions.

For clarity, the datasets employed are following: BallRoom [3], CretanDances, BRID [5], Candombè [8, 7], and MalianJembe [9].

Click here to visualize the interactive plot generated from the first clustering experiment and click here to visualize the interactive plot generated from the second clustering experiment.

Overall, these two experiments served to demonstrate the application of the scale transform magnitude.

3.2 Computation and Transformation

I implemented one notebook to demonstrate all the individual steps for the computation of the STM. This serves an educational purpose, helping to understand the underlying computation of the feature in more detail.

Additionally, I created another notebook to showcase the robustness of the feature to changes in speed, timbre, and dynamics. This involved applying certain types of transformations to the signal, such as time stretching, and then computing the cosine similarity between the original version and the transformed version of the signal. Results showed a very high cosine similarity, proving the scale-invariance property of the scale transform magnitude.

4 Conclusion

During this internship, I familiarized myself with the available tools and data to conduct research on rhythmic similarity, and I deepened my knowledge of digital signal processing. The repository containing all the code produced during the last few months can be found here.

References

[1] Antonio De Sena and Davide Rocchesso. A fast mellin and scale transform. *EURASIP Journal on Advances in Signal Processing*, 2007:1–9, 2007.

[2] Tlacael Miguel Esparza, Juan Pablo Bello, and Eric J Humphrey. From genre classification to rhythm similarity: Computational and musicological insights. *Journal of New Music Research*, 44(1):39–57, 2015.

[3] Fabien Gouyon, Anssi Klapuri, Simon Dixon, Miguel Alonso, George Tzanetakis, Christian Uhle, and Pedro Cano. An experimental comparison of audio tempo induction algorithms. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5):1832–1844, 2006.

[4] André Holzapfel and Yannis Stylianou. Scale transform in rhythmic similarity of music. *IEEE transactions on audio, speech, and language processing*, 19(1):176–185, 2010.

[5] Lucas S. Maia, Pedro D. de Tomaz Júnior, Magdalena Fuentes, Martín Rocamora, Luiz W. P. Biscainho, Maurício V. M. Costa, and Sara Cohen. A novel dataset of brazilian rhythmic instruments and some experiments in computational rhythm analysis. 2018.

[6] Lucas Simões Maia, Martín Rocamora, Luiz WP Biscainho, and Magdalena Fuentes. Selective annotation of few data for beat tracking of latin american music using rhythmic features. *Transactions of the International Society for Music Information Retrieval, vol. 7, no 1, may 2024, pp. 99-112.*, 2024.

[7] Luis Jure Martın Rocamora, Bernardo Marenco, Magdalena Fuentes, Florencia Lanzaro, and Alvaro Gómez. An audio-visual database of candombe performances for

computational musicological studies. In *Proceedings of the II Congreso Internacional de Ciencia y Tecnologıa Musical*, pages 17–19, 2015.

[8] Leonardo Nunes, Martín Rocamora, Luis Jure, and Luiz 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 (ISMIR 2015)*, pages 264–270, Málaga, Spain, Oct. 2015.

[9] Rainer Polak, Simone Tarsitani, and Martin Clayton. Iemp malian jembe. *Open Science Framework*, 2018.

[10] WJ Williams and EJ Zalubas. Helicopter transmission fault detection via time-frequency, scale and spectral methods. *Mechanical systems and signal processing*, 14(4):545–559, 2000.