MÖBIUS BEATS: THE TWISTED SPACES OF SLIDING WINDOW AUDIO NOVELTY FUNCTIONS WITH RHYTHMIC SUBDIVISIONS

Christopher J. Tralie
Duke University
Department of Mathematics

John Harer
Duke University
Department of Mathematics

ABSTRACT

In this work, we show that the sliding window embeddings of certain audio novelty functions (ANFs) representing songs with rhythmic subdivisions concentrate on the boundary of non-orientable surfaces such as the Möbius strip. This insight provides a radically different topological approach to classifying types of rhythm hierarchies. In particular, we use tools from topological data analysis (TDA) to detect subdivisions, and we use thresholds derived from TDA to build graphs at different scales. The Laplacian eigenvectors of these graphs contain information which can be used to estimate tempos of the subdivisions. We show a proof of concept example on an audio snippet from the MIREX tempo training dataset, and we hope in future work to find a place for this in other MIR pipelines.

1. INTRODUCTION

Automatic rhythm understanding in audio is a long standing problem in MIR. Most techniques for rhythm analysis start with an audio novelty functions (ANFs), which are a downsampled version of the original audio signal meant to correlate with rhythmic events, and which are usually derived from spectrograms [2, 3, 9, 12]. Most approaches to beat tracking and tempo analysis use dynamic programming [9, 14], a Bayesian approach [6, 22], or some type of autocorrelation [7, 16], Fourier [16, 18], or wavelet [21] analysis. We take a completely orthogonal approach by considering a geometric, dynamical systems perspective on ANFs. The end result is a pipeline in which shape properties of sliding window embeddings ANFs can be used to uncover ratios between different rhythm levels in audio.

2. SLIDING WINDOW EMBEDDINGS OF PULSES

Given $M$ lags and an interval $\tau$, the sliding window embedding of a 1D function $f(t)$ \footnote{For ease of exposition, we define the sliding window as acting on continuous 1D functions, but in practice these functions are discretized to $N$ samples, and interpolation may be necessary for some $M, \tau$ choices.} is the space curve parameterized as

$$S_{M, \tau}[f](t) = \begin{bmatrix} f(t) \\ f(t + \tau) \\ \vdots \\ f(t + M\tau) \end{bmatrix} \in \mathbb{R}^{M+1} \tag{1}$$

Under the right conditions, sliding window embeddings of time series which witness deterministic processes can be used to reconstruct the state spaces of those processes [13, 20]. It is for this reason that the authors in [19] advocated for sliding windows of Chroma vectors as a preprocessing step to improve robustness. A simpler example is that of a pure sinusoid, for which $S_{M, \tau}\{e^{i2\pi t}\} = u e^{i2\pi t} + v \sin(t)$ for two fixed vectors $u, v \in \mathbb{R}^{M+1}$ (see [17]), which is an equation parameterizing an ellipse. More generally, as shown by the authors of [17], the sliding window embedding of any periodic function (i.e. $f(t) = f(t + T)$ for some $T \in \mathbb{R}^+$) lies on a topological loop, though the geometry may be quite complicated. For instance, the sliding window embedding of $a\cos(t) + b\cos(2t)$ lies on the boundary of a Möbius strip if $b > a$ [17] (note that the boundary of a Möbius strip is a single loop, see Figure 2). Inspired by this result, we inves-
SSMs for smoothed pulses for different cases. For a system with windows in time have a finite distance. Figure 1 shows the pairwise distances between windows. This is a combinatorial object with a vertex for each class, with its birth time on the x-axis and death time on the y-axis. At a scale equal to the strip width $g$, another class is born, which dies at the maximum distance $m$. By contrast, for all other field coefficients, there is only one significant class which is born at $r$ and dies at $m$ (see [17] for a similar example with pure sinosoids). In general, for finite fields with $p$ elements, where $p$ is a prime factor of $k$, this “splitting” of one class $[r, m]$ into $[r, g]$ and $[g, m]$ will occur, which can be used to identify subdivision. Figure 4 shows a real 3 on 1 example using the audio novelty function from [9].

2.2 Graph Laplacian Circular Coordinates

We now turn to spectral graph theory to help uncover tempos of the different subdivisions, taking inspiration from [1]. Let $A$ be the adjacency matrix of a graph, and let $D$ be the degree matrix. Then $L = D - A$ is the unweighted graph Laplacian. We can build a graph on the discrete set of windows. As hinted at in the SSMs (Figure 1), if we include edges with distances under the birth times in the persistence diagrams $r$ and $g$, then we always end up with “circulant graphs,” or graphs in which $A$ is circulant [11], which have Laplacians diagonalized by the Discrete Fourier Transform. For the sliding window of a pulse train with period $T$ and subdivision by factor $k$, if we only include edges up to window neighbor threshold $r$ in the graph, we get a loop graph. The eigenvectors $L$ with the smallest two nonzero eigenvalues are orthogonal linear combinations of $\cos(2\pi n/T), \sin(2\pi n/T)$. In the case that we include edges every $k$ lags (threshold $g$), the eigenvectors with the smallest two nonzero eigenvalues are $\cos(2\pi kn/T)$ and $\sin(2\pi kn/T)$. The absolute slope of $\theta[n] = \tan^{-1}(v_2[n]/v_1[n])$ gives a tempo at each scale.

3 By scale $x$, we mean a “Rips complex” built from distance information between windows. This is a combinatorial object with a vertex for each window, edges between windows that are at most $x$ apart, and triangles between triples of windows which are pairwise at most $x$ apart.

4 $A_{ij} = 1$ if edge from $i$ to $j$, or 0 otherwise. $D_{ii} = \sum_{j=1}^N A_{ij}$
3. REFERENCES


