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 ANFns can be used to uncover ratios between different rhythm levels in audio.

2. SLIDING WINDOW EMBEDDINGS OF PULSES

Given M lags and an interval τ, the sliding window embedding of a 1D function $f(t)$ [1] is the space curve is parameterized as

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

[1] 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, τ choices.

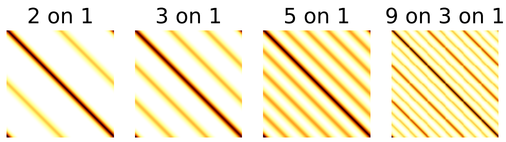

Figure 1. Self-similarity matrices (SSMs) of sliding window embeddings of various harmonic pulse trains.

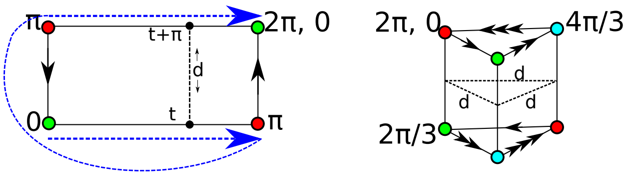

Figure 2. (Left) The Möbius strip (2 on 1) with its boundary drawn in blue and arrows showing identifications (glued locations). In this rendering, the boundary jumps from the lower right corner to the upper right corner of this diagram because of the twist. A uniform Möbius strip has the property that for its boundary $[0, 2\pi] \rightarrow \boldsymbol{X(t)}$, $||\boldsymbol{X}(t) - \boldsymbol{X}(t+\pi)||_2$ is a constant d (the width of the strip). (Right) Analogous structure for the 3 on 1 geometry.

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 $\boldsymbol{S}_{M,\tau}[\cos](t) = \boldsymbol{u}\cos(t) + \boldsymbol{v}\sin(t)$ for two fixed vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{M+1}$ (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-

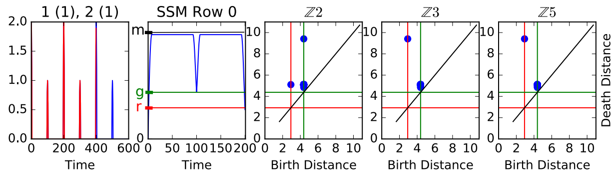

Figure 3. Sliding window of the 2 on 1 pulse train, $T = 200$, $k = 2$, $a = b = 1$, $M\tau = 2T$, and the first window is highlighted in red on the left plot. The horizontal red line indicates the distance between adjacent windows in time, and the green line indicates the local min distance at $T/2$. Persistence diagrams for fields $\mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_5$ are on the right.

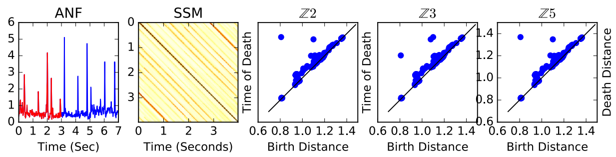

Figure 4. Sliding window embedding of the audio novelty function (ANF) from part of Erykah Badu's "Green Eyes," which has a 3 on 1 structure, as indicated by the $\mathbb{Z}_3$ change.

tigate the sliding window embeddings of "k on 1 harmonic pulse trains" as a model for ANFs with subdivision; that is

$$f(t) = a\delta(t(\text{mod}T)) + b\delta(kt(\text{mod}T)) \quad (2)$$

for some constants a, b and some positive integer k, where δ is the Kronecker delta. Given $M\tau = 2T$, we observe that the pairwise distances between windows are

$$d(s,t) = \left\{ \begin{array}{cc} 0 & |s-t| = lT \\ 2|b-a| & |s-t| = l'T + T/k \\ \infty & \text{otherwise} \end{array} \right\} \quad (3)$$

for $l, l' \in \mathbb{Z}$. That is, the windows line up perfectly after an integer number of periods, but they also line up locally at every subdivided beat shift. Otherwise, they don't line up at all, though in practice we smooth the pulses so adjacent windows in time have a finite distance. Figure 1 shows SSMs for smoothed pulses for different cases. For $k = 2$, this matches the geometry of the Möbius strip boundary, which is locally close to itself at a lag of $T/2$ (Figure 2).

2.1 Discovering Subdivisions with TDA

To discover these shapes in data, we use techniques from topological data analysis (TDA). In broad strokes [2], TDA provides a way to quantify "cycles" (connected components, loops, voids) at different scales in point cloud data. It uses a computable invariant known as *homology*, which turns the problem of quantifying these features into a linear algebra problem. For certain shapes with "torsion," such as the Möbius strip, the field of coefficients used in the vector space representing the objects can change the homology [15]. In the case of the Möbius strip boundary, if we use $\mathbb{Z}_2$ (binary) coefficients, a loop class is "born" (i.e.

[2] This is an intricate subject, details are beyond our scope; see [4,8,10]

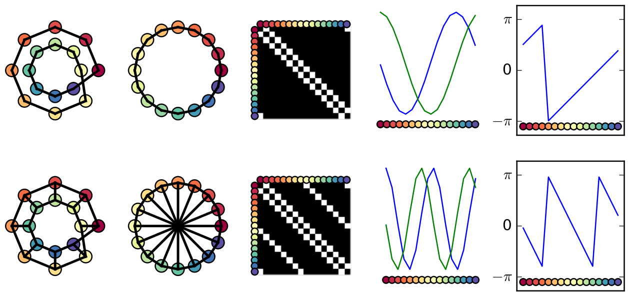

Figure 5. Different thresholds on the Möbius ladder graph. Columns 1+2 show different layouts of the graph, column 3 shows the corresponding adjacency matrices, column 4 shows the first two nonzero eigenvectors v_1 (blue) and v_2 (green), and the last column shows $\tan^{-1}(v2/v1)$.

forms for the first time) at a scale [3] r equal to the distance between adjacent windows (red line, Figure 3) and "dies" (i.e. fills in) at a scale slightly larger than the strip width g (green line). These changes are summarized in a "persistence diagram" which has a dot for every 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 sinusoids). 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 [5] 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 [4]. 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 $\cos(2\pi kn/T)$. The absolute slope of $\theta[n] = \tan^{-1}(\frac{v_2[n]}{v_1[n]})$ 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

[1] Hadar Averbuch-Elor and Daniel Cohen-Or. Ringit: Ring-ordering casual photos of a temporal event. *ACM Trans. Graph.*, 34(3):33–1, 2015.

[2] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Chris Duxbury, Mike Davies, and Mark B Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on speech and audio processing*, 13(5):1035–1047, 2005.

[3] Sebastian Böck and Gerhard Widmer. Maximum filter vibrato suppression for onset detection. In *Proc. of the 16th Int. Conf. on Digital Audio Effects (DAFx). Maynooth, Ireland (Sept 2013)*, 2013.

[4] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[5] Fan RK Chung. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.

[6] Norberto Degara, Enrique Argones Rúa, Antonio Pena, Soledad Torres-Guijarro, Matthew EP Davies, and Mark D Plumbley. Reliability-informed beat tracking of musical signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):290–301, 2012.

[7] Douglas Eck. Beat tracking using an autocorrelation phase matrix. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 4, pages IV–1313. IEEE, 2007.

[8] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[9] Daniel PW Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007.

[10] Robert W Ghrist. *Elementary applied topology*. Createspace, 2014.

[11] Chris Godsil and Gordon F Royle. *Algebraic graph theory*, volume 207. Springer Science & Business Media, 2013.

[12] Fabien Gouyon, Simon Dixon, and Gerhard Widmer. Evaluating low-level features for beat classification and tracking. In *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, volume 4, pages IV–1309. IEEE, 2007.

[13] Holger Kantz and Thomas Schreiber. *Nonlinear time series analysis*, volume 7. Cambridge university press, 2004.

[14] Jean Laroche. Efficient tempo and beat tracking in audio recordings. *Journal of the Audio Engineering Society*, 51(4):226–233, 2003.

[15] James R Munkres. *Elements of algebraic topology*, volume 2. Addison-Wesley Menlo Park, 1984.

[16] Geoffroy Peeters. Time variable tempo detection and beat marking. In *ICMC*, pages 539–542, 2005.

[17] Jose A Perea and John Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15(3):799–838, 2015.

[18] Elio Quinton, Christopher Harte, and Mark Sandler. Extraction of metrical structure from music recordings. In *Proc. of the 18th Int. Conference on Digital Audio Effects (DAFx). Trondheim*, 2015.

[19] Joan Serra, Xavier Serra, and Ralph G Andrzejak. Cross recurrence quantification for cover song identification. *New Journal of Physics*, 11(9):093017, 2009.

[20] Floris Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.

[21] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302, 2002.

[22] Nick Whiteley, Ali Taylan Cemgil, and Simon J Godsill. Bayesian modelling of temporal structure in musical audio. In *ISMIR*, pages 29–34, 2006.