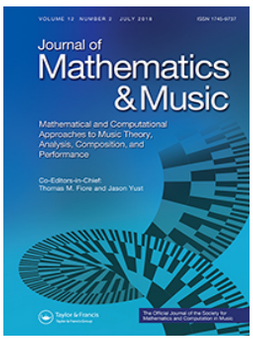


Taylor & Francis
Taylor & Francis Group

Journal of Mathematics and Music
Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tmam20

Homological persistence in time series: an application to music classification

Mattia G. Bergomi & Adriano Baratè

To cite this article: Mattia G. Bergomi & Adriano Baratè (2020): Homological persistence in time series: an application to music classification, Journal of Mathematics and Music, DOI: 10.1080/17459737.2020.1786745

To link to this article: https://doi.org/10.1080/17459737.2020.1786745

Published online: 14 Jul 2020.

Submit your article to this journal

Article views: 53

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tmam20

Journal of Mathematics and Music, 2020
https://doi.org/10.1080/17459737.2020.1786745

Taylor & Francis
Taylor & Francis Group

Check for updates

Homological persistence in time series: an application to music classification

Mattia G. Bergomi [a*] and Adriano Baratè[b]

[a]*Veos Digital, Milan, Italy;* [b]*Music and Computer Science Laboratory, University of Milan, Milan, Italy*

(*Received 22 May 2019; accepted 2 June 2020*)

Meaningful low-dimensional representations of dynamical processes are essential to better understand the mechanisms underlying complex systems, from music composition to learning in both biological and artificial intelligence. We suggest to describe time-varying systems by considering the evolution of their geometrical and topological properties in time, by using a method based on persistent homology. In the static case, persistent homology allows one to provide a representation of a manifold paired with a continuous function as a collection of multisets of points and lines called persistence diagrams. The idea is to fingerprint the change of a variable-geometry space as a time series of persistence diagrams, and afterwards compare such time series by using dynamic time warping. As an application, we express some music features and their time dependency by updating the values of a function defined on a polyhedral surface, called the *Tonnetz*. Thereafter, we use this time-based representation to automatically classify three collections of compositions according to their style, and discuss the optimal time-granularity for the analysis of different musical genres.

Keywords: *Tonnetz*; topology; time-series analysis; persistent homology; dynamic time warping; classification; style

2010 Mathematics Subject Classification: 00A65; 20B35; 18B25
2012 Computing Classification Scheme: Applied computing

1. Introduction

A time-varying system can be interpreted as a series of relevant geometric and topological events. Persistent homology has mainly been applied to the study of static point clouds and shapes (Edelsbrunner and Harer 2009), by providing a description of both the geometry and topology of the analyzed space. One of the reasons that make persistent homology so effective when applied to the study of static spaces is that it provides a representation in which the features of the space appear arranged by relevance. Thus, the analysis can be tuned on a specific application need, by balancing the computational cost and the retrieval of details. Briefly one can think about persistent homology as a scalable fingerprint of a static object.

Persistent homology has been generalized to time-varying systems either by considering continuous representations (Cohen-Steiner, Edelsbrunner, and Morozov 2006), or introducing statistics, in order to evaluate the evolution in time of the analyzed system (Munch 2013; Turner et al. 2014).

*Corresponding author. Email: mattia.bergomi@veos.digital

© 2020 Informa UK Limited, trading as Taylor & Francis Group

A time series is a collection of values obtained through subsequent observations in time. Time-series data mining (Esling and Agon 2012) is an attempt to organize data, in furtherance of visualizing the *contour* of data, avoiding negligible details and creating a consistent, interpretable representation. Due to their generality and flexibility, time series are extensively used in applications, e.g. classification (Bakshi and Stephanopoulos 1994), segmentation (Keogh et al. 2004), and supported by a strong theoretical framework (Keogh and Kasetty 2003).

We suggest to combine the scalability (among other properties) of persistent homology and the notion of time series, in order to characterize the evolution of a variable-geometry space in time. We will discuss how this representation allows one to distinguish between relevant and noisy observations by utilizing the natural notion of distance defined in the context of persistent homology as a cost function (Senin 2008). In addition, the computation of the dissimilarity between time series (Liao 2005) will allow us to compare several time-varying systems; i.e. to find the timespans (if they exist) where two time series can be considered comparable in agreement with their time-dependent geometric and topological characterization.

As an application, we consider a variable-geometry polyhedral surface T inspired by the *Tonnetz*, a graph widely used in computational musicology to represent the relationships among notes. We will perform automatic stylistic clustering of three collections of classical, jazz, and pop songs, by updating the geometry of T according to the information deduced from a certain composition, and therefore computing the associated time series of persistence diagrams.

2. Persistent homology background

In computational topology (Edelsbrunner and Harer 2009), persistent homology provides a multiscale description of the topology of a given space. The efficacy and versatility of this technique is demonstrated by its numerous and diverse applications. See Ferri (2017) for a review. In the remainder of this section, we will provide a basic intuition of homology and persistent homology.

2.1. *Homology*

Homology is a standard theory in Algebraic Topology first described in Poincaré (1895). Here we will give an intuition about homology and limit ourselves to discuss the basic concepts and definitions needed in the remainder of this work. However, homology can be defined in more general settings. We refer the reader to Munkres (1984) and Hatcher (2002) for a detailed, rigorous description.

The extremely simplified view on homology needed for the understanding of what follows is that it allows one to map a topological space X (think of a manifold for simplicity) to a sequence of algebraic objects, here vector spaces denoted by $H_i(X)$ for every $i \in \mathbb{Z}$. Although $H_i(X)$ are not always vector spaces, it is possible to achieve this structure by computing homology with coefficients in a field. We compute homology in $\mathbb{Z}/2\mathbb{Z}$. See Munkres (1984), Hatcher (2002) and Bergomi (2015) (Part 1, Chapter 2), for more details. The dimension of the vector space $H_i(X)$ counts the number of i-dimensional *holes* of X. Thus, $\dim(H_0(X))$ corresponds to the number of connected components of X, the dimension of $H_1(X)$ to the number of holes, $H_2(X)$ voids, and so on. We define the i-Betti number associated with X as $\beta_i(X) = \dim(H_i(X))$, or simply β_i when no confusion arises. See Figure 1 for an example.

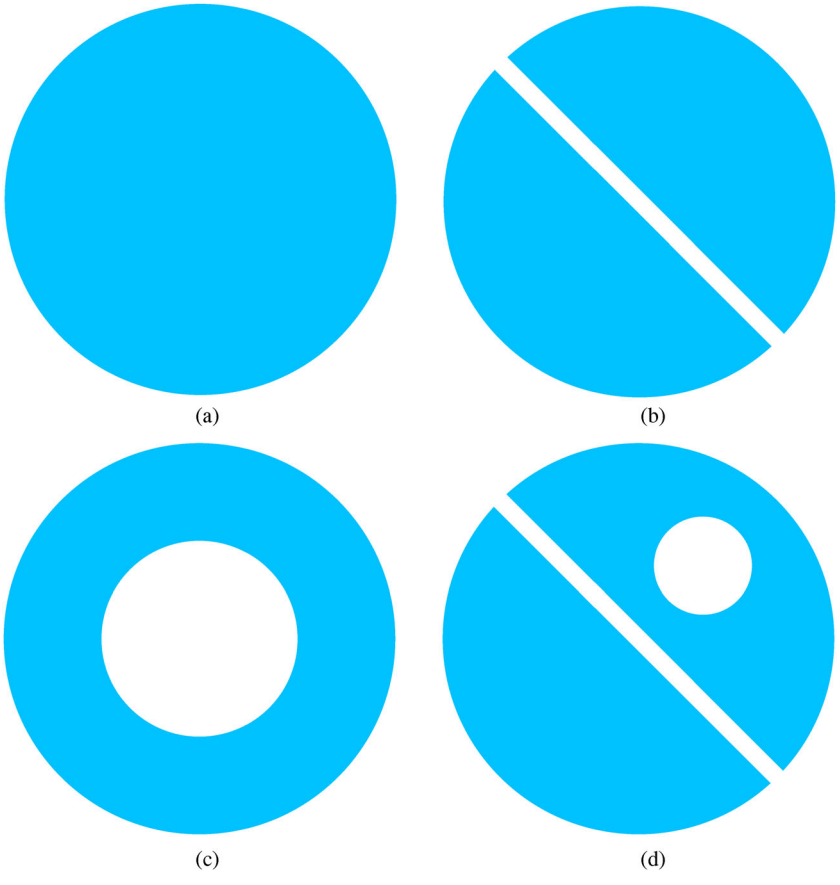


Figure 1. An intuition on Betti numbers. For each of the four example spaces, the 0-Betti number β_0 counts the number of connected components, while β_1 counts 1-dimensional holes. (a) $\beta_0 = 1, \beta_1 = 0$. (b) $\beta_0 = 2, \beta_1 = 0$. (c) $\beta_0 = 1, \beta_1 = 1$ and (d) $\beta_0 = 2, \beta_1 = 1$.

2.2. *Persistent homology*

Clearly, the computation of the Betti numbers associated with a certain space X causes a huge loss of information. Persistent homology allows us not only to alleviate, but control this phenomenon, by considering the evolution of the Betti numbers on a *filtration* of X, i.e. a sequence of nested spaces starting with the empty set and ending with the entire space X.

2.2.1. *Filtering functions*

Let X be a triangulable manifold and $f : X \to \mathbb{R}$ a continuous function. A filtration of X is defined as a collection of nested sublevel sets $X_u = f^{-1}((-\infty, u]),\ u \in \mathbb{R}$. The function f is called a *filtering function* or simply a *filter*.

The choice of f depends on the geometric property of X one wants to take into account. To understand how the choice of a filtering function alters the representation of a given space, it is necessary to introduce the following definition. A real number r is a *regular homological value* for the pair (X, f) if it exists $\varepsilon > 0$ such that for every pair of real numbers $x < y$ in $(r - \varepsilon, r + \varepsilon)$, the homology of X_x is isomorphic to the one of X_y. For the purposes of this manuscript, one can simply think that X_x and X_y have the same Betti numbers. Otherwise r is said to be a homological

critical value of (X, f). See Govc (2013) for details. A filtering function with a finite number of homological critical values $\{a_1, \ldots, a_n\}$, finite i-Betti numbers for each sublevel set X_{a_k}, $k \in \{1, \ldots, n\}$ and $i \in \mathbb{Z}$ is said to be *tame*.

We can now consider the filtration

$$\emptyset \subseteq X_{a_1} \subseteq X_{a_2} \subseteq \cdots \subseteq X_{a_n} = X,$$

compute the Betti numbers associated with each sublevel set, and observe their evolution throughout the filtration. The scale at which a feature is significant is measured by considering its longevity (i.e. persistence) along the filtration.

2.2.2. *Persistence diagrams*

All the information carried by this analysis can be summarised in a simple and computationally efficient form. We do this by enumerating the values of birth and death of each component that produces a change in the Betti numbers along the filtration. Let us denote each of these points by (b_i, d_i). The multiset (a set whose elements can be repeated) $D = \{(b_i, d_i)\}_{i \in \{1,\ldots,m\}}$ is called *persistence diagram*. Graphically, a persistence diagram is represented as a collection of points in $\mathbb{R}^2$ with the following characteristics:

(1) $d_i \geq b_i$ for every $(b_i, d_i) \in D$. This condition means that a n-dimensional hole cannot die before it is born.
(2) The ordinate d_i can be infinite. It is indeed possible that a n-dimensional hole that is born along the filtration never dies.

The points of the persistence diagram are called *cornerpoints*; points of the form (b_i, ∞) with infinite lifespan are said to be *cornerlines* and graphically rendered as vertical half lines. The distance of a cornerpoint from the diagonal is its *persistence*. See Figure 2(a,b) for an example. In the figure, observe how the local minima a_2 and a_4, when merging with the maxima a_3 and a_5 give rise to cornerpoints with different persistence, because $a_3 - a_2 < a_5 - a_4$. Finally, observe that more cornerpoints can have the same coordinates. This generally reflects the presence of symmetries of the analyzed space with respect to the filtering function. Set of overlapping cornerpoints are considered as a single cornerpoint with multiplicity equal to the cardinality of the set.

2.2.3. *Bottleneck distance*

The *bottleneck distance* (d'Amico, Frosini, and Landi 2006) is naturally a metric between persistence diagrams defined as

$$d_B(D, D') = \min_{\sigma} \max_{p \in D} d(p, \sigma(p)),$$

where D and D' are two kth persistence diagrams, σ varies among all the bijections between D and D', and

$$d\left((u, v), \left(u', v'\right)\right) = \min\left\{\max\left\{|u - u'|, |v - v'|\right\}, \max\left\{\frac{v - u}{2}, \frac{v' - u'}{2}\right\}\right\},$$

for every $(u, v), (u', v') \in \{(x, y) \in \mathbb{R}^2 : x \leq y\}$.

The bottleneck distance is the most used metric to compare persistence diagrams, because of its properties of optimality (d'Amico, Frosini, and Landi 2010) and stability (Cohen-Steiner,

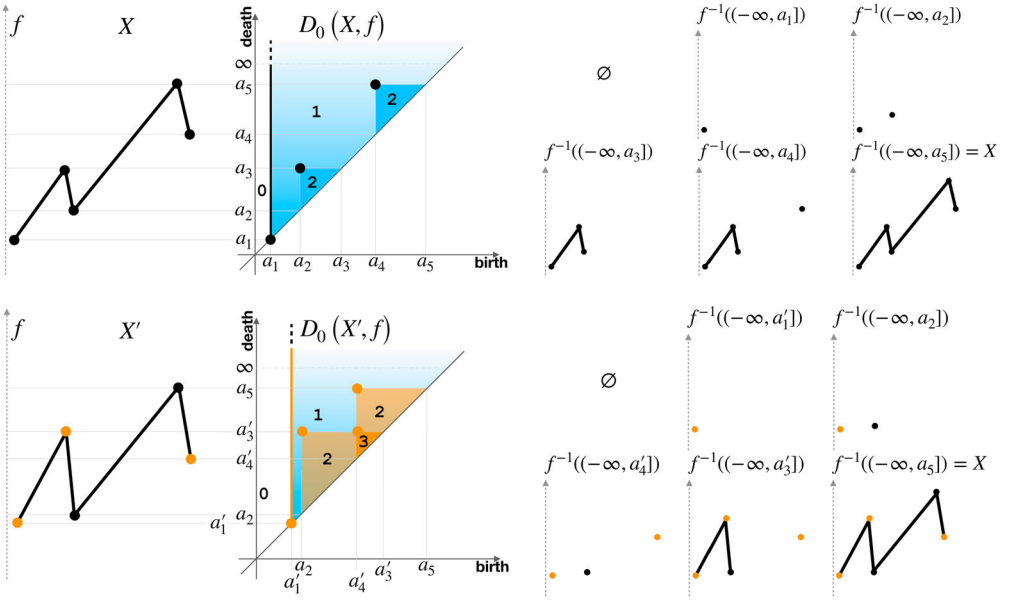


Figure 2. Panel (a): persistent homology induced by the sublevel sets of the height function f on the space X. $\{a_i\}_{i\in\{1,\dots,5\}}$ are the ordinates of local maxima and minima of $f : X \to \mathbb{R}$. The 0-persistence diagram on the right of the panel shows the evolution of the 0-Betti number (number of connected components) throughout the filtration $f^{-1}((-\infty, a_i])$. The integer numbers reported in the persistence diagram correspond to the value of the 0-persistent Betti number function for every $(u, v) \in \mathbb{R}^2$. For instance, before the sublevel set $f^{-1}((-\infty, a_1])$ is reached, no connected component exists. For values $a_2 \leq f(x) < a_3$ two connected components exist. Those will merge in a single connected component once a_3 is reached. The cornerline of coordinates (a_1, ∞) is associated with the connected component corresponding to the absolute minimum of f with value a_1. Each step of the filtration is depicted in Panel (b). Panel (c): a variation X' (in orange) of the space X of Panel (a). The variation is detected by the 0-persistence diagram and highlighted in orange. The filtration associated with (X', f) is illustrated in Panel (d). (a) 0-persistence diagram. (b) Sublevel set filtration of X with respect to f. (c) A variation X' of the space X and its 0-persistence diagram and (d) Sublevel set filtration of X' with respect to f.

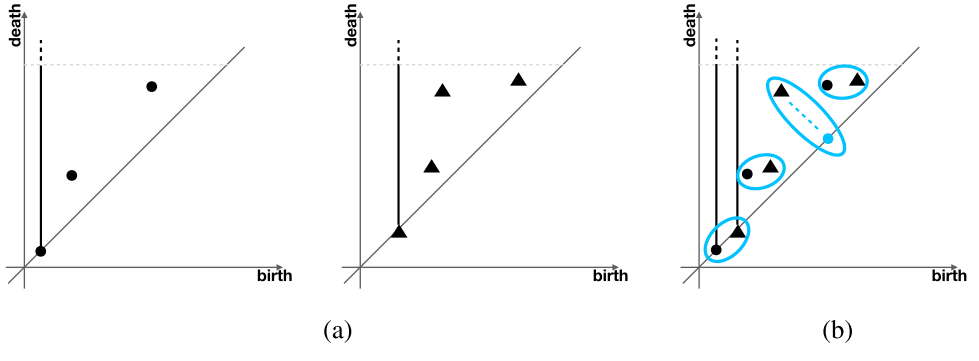
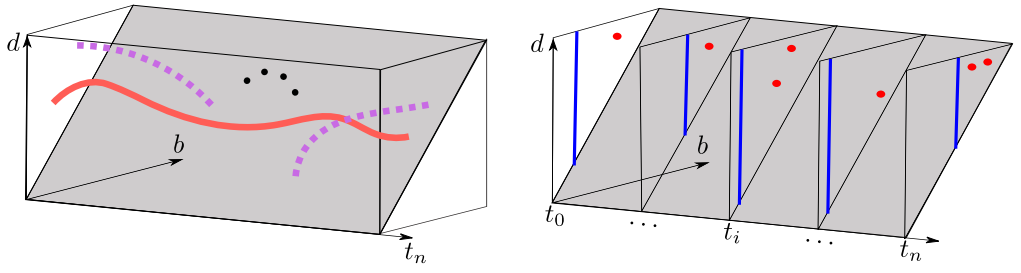


Figure 3. (a) Two persistence diagrams. (b) Cornerpoints are matched by choosing to the bijection that minimises the sum of the distances. Observe how one of the cornerpoints represented as triangles is matched with its projection on the diagonal. (a) Persistence diagrams and (b) Matching.

Edelsbrunner, and Harer 2007). See Figure 3 for an intuition. It should be clear from the figure that in order to measure the distance between two persistence diagrams with a different number of cornerpoints, it is necessary to consider bijections that map some cornerpoints placed on the diagonal, i.e. with zero persistence.



(a) (b)

Figure 4. (Left) An example of vineyard. The axes represent the time t, the birth level b and the death level d of the k-homology classes of a system evolving in time. Vines are represented as continuous paths. (Right) Persistence diagrams associated with evenly spaced observations of a time-varying system.

2.3. *Time-varying systems*

A change of the geometry of a shape corresponds to an update of its persistence diagrams. This update encodes the relevance of change in time, with respect to the filtering function. See Figure 2.

Persistent homology has been generalized to the study of time-varying systems in Cohen-Steiner, Edelsbrunner, and Morozov (2006). Let $f, g : \mathbb{X} \to \mathbb{R}$ be tame functions, $I = [0,1] \subset \mathbb{R}$ and $H : \mathbb{X} \times I \to \mathbb{R}$ a homotopy between f and g, i.e. a continuous deformation between f and g. See Hatcher (2002) for more details on homotopy theory. Assuming that $H(x,t)$ is tame for every $t \in I$, we obtain a k-persistence diagram for every pair $(t,k) \in I \times \mathbb{Z}$, named a *vineyard*. In this representation, each cornerpoint is associated with a trajectory called a *vine* (cornerlines can be seen as cornerpoints (u,v), such that $v \gg u$). See Figure 4 for an example.

3. Persistence time series

Vineyards are powerful tools for describing time-varying systems. Indeed, given a time-varying system (X,t) they encode the changes in time of the homological critical values induced by a given filtering function. However, their interpretation is not intuitive, and there is not a general technique that allows one to compare vineyards associated with two time-varying systems.

Let $\mathbb{X}$ be a topological space, $f : \mathbb{X} \times I \to \mathbb{R}$ a tame function for every $t \in I$, and $\mathbf{t} = \{ t_0, \ldots, t_n \}$ a set of $(n+1)$ evenly spaced points of I. A k-persistence diagram $D_k(X,f)_i$ is associated with every $t_i \in \mathbf{t}$. The collection of these *k-persistence snapshots* is a time series $\mathcal{D}_{k,n}(X,f) = \{ D_k(X,f)_i \}_{i=0}^{n} \subset D_\infty$, where (D_∞, d_B) is the space of persistence diagrams equipped with the bottleneck distance. We name $\mathcal{D}_{k,n}(X,f)$, or simply $\mathcal{D}(X,f)$, a *k-persistence time series*.

There exist several methods to evaluate the dissimilarity of two time series (Liao 2005; Esling and Agon 2012; Lines and Bagnall 2015). Here, we will consider Dynamic Time Warping (DTW) (Berndt and Clifford 1994; Senin 2008). One of the most important features of DTW is that it allows us to align, and hence measure the dissimilarity between time series of different lengths, by providing their global optimal warping.

3.1. *Dynamic time warping*

Let $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ be two time series, where $n, m \in \mathbb{N}$. The first ingredient needed for the computation of the optimal warping path between X and Y is a measure of

dissimilarity between pairs of observations. The DTW algorithm will align the two time series by rearranging their points, in order to minimise their global dissimilarity.

Let us assume that the observations of both time series belong to a space Φ, said to be the *feature space*. The dissimilarity function, or *cost function*, is defined as $c : \Phi \times \Phi \to \mathbb{R}$, such that

(1) $c(x, y) \geq 0$ for all observations $x, y \in \Phi$;
(2) $c(x, y) = 0$ if and only if $x = y$.

Observe that a cost function is not necessarily a metric.

The second step is the construction of the *local cost matrix* $C \in \mathbb{R}^{n\times m}$, whose entries are defined as $C_{ij} := c(x_i, y_j)$, for $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$. A warping path $\gamma = (\gamma_1, \ldots, \gamma_l)$ is a sequence of pairs of indices $\gamma_k = (i_k, j_k)$, where $i_k \in \{1, \ldots, n\}$, $j_k \in \{1, \ldots, m\}$ and $k \in \{1, \ldots, l\}$, such that the following conditions hold:

(1) $\gamma_1 = (1, 1)$ and $\gamma_l = (m, n)$ for every $l \in \mathbb{N}$. This condition simply states that the starting and ending observations of the two time series have to be aligned, i.e. we are aligning the two time series globally.
(2) Given $\gamma_k = (i_k, j_k)$ and $\gamma_{k+1} = (i_{k+1}, j_{k+1})$, we have $i_k \leq i_{k+1}$ and $j_k \leq j_{k+1}$. This condition ensures that the natural ordering induced by time on the observations is preserved.
(3) The difference between two subsequent steps in the warping path $\gamma_{k+1} - \gamma_k \in \{(0, 0), (1, 0), (0, 1), (1, 1)\}$. This last condition constrains the size of the shifts in time used to align the two time series.

The *total cost* of a (n, m)-warping path γ over the observations of X and Y is defined as

$$c_\gamma(X, Y) := \sum_{k=1}^{l} c(x_{i_k}, y_{j_k}).$$

An *optimal warping path* on X and Y is a warping path realising the minimum total cost. We are now ready to define the *DTW distance* between X and Y:

$$DTW(X, Y) := \min \left\{ c_\gamma(X, Y) \;\middle|\; \gamma \text{ is a } \right\} (n, m)\text{-warping path.}$$

Note that the minimum always exists because the set is finite. In Figure 5, a simple example of global alignment of time series *via* DTW is depicted.

3.2. *DTW for persistence time series*

When considering persistence time series, the bottleneck distance can be used as a cost function. Let $\mathcal{D}_{k,n}(X, f)$ and $\mathcal{D}_{k,m}(Y, g)$ be two kth persistence time series. By definition, the bottleneck distance satisfies the properties that characterize a cost function. The DTW between k-persistence time series is given by an optimal warping path defined as:

$$\begin{aligned} &\text{DTW}\left(\mathcal{D}(X, f), \mathcal{D}(Y, g)\right) \\ &\quad = \min \left\{ d_{B_\gamma}\left(\mathcal{D}(X, f), \mathcal{D}(Y, g)\right) \;\middle|\; \gamma \text{ is an } (n, m) - \text{ warping path} \right\}. \end{aligned}$$

The DTW inherits the symmetry by the bottleneck distance, and the tameness of f and g assures stability of every diagram $D(X, f_i)$ and $D(Y, g_j)$, for every $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$.

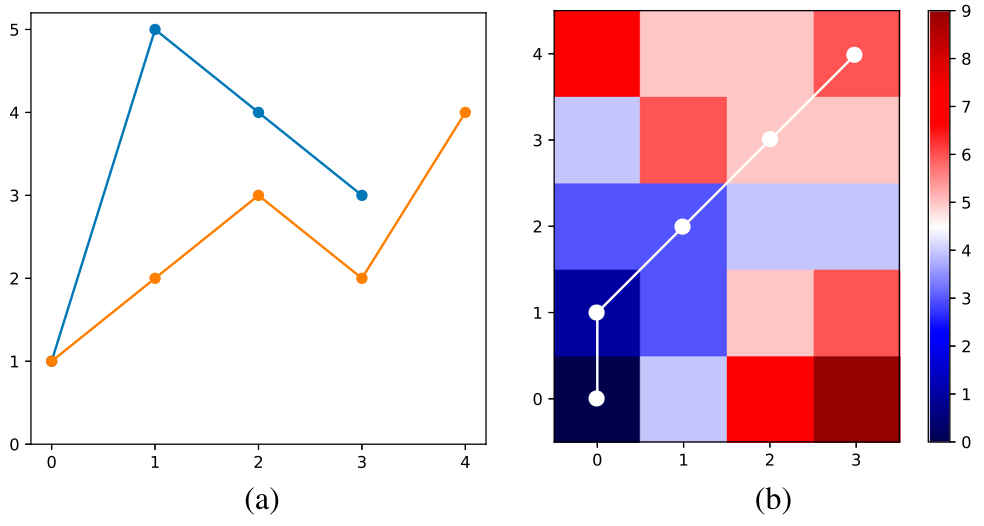


Figure 5. (a) Two time series of length 5 and 4, respectively. (b) The cost matrix and optimal warping path between the two time series.

Consider the prefix sequences $\mathcal{D}_{k,\alpha}(X,f)$ and $\mathcal{D}_{k,\beta}(Y,g)$, where $\alpha \leq n$ and $\beta \leq m$ are natural numbers. The *accumulated cost matrix* is defined as

$$A(\alpha,\beta) := DTW\left(\mathcal{D}_{k,\alpha}(X,f), \mathcal{D}_{k,\beta}(Y,g)\right).$$

Then $A(n,m) = DTW\left(\mathcal{D}_{k,n}(X,f), \mathcal{D}_{k,m}(Y,g)\right)$.

4. Musical setting

Music, or at least some of its features, can be represented using a variable-geometry space. Before illustrating our model, it is necessary to introduce some basic definitions.

A *note in equal tuning*, denoted by $\mathfrak{n}$, can be modeled as a pair $(p,d) \in \mathbb{R}^2$, where p is its *pitch*, and d its *duration* in seconds. In particular, if ν denotes the fundamental frequency of $\mathfrak{n}$, we have $p(\nu) = 69 + 12\log_2(\nu/440)$, where 440 Hz is the fundamental frequency of the A of the fourth octave of the piano (Cheveignéde 2005).

Perceptively, two notes an octave apart are really similar (Burns and Ward 1999), thus, it is a common practice in computational musicology to identify pitches modulo octave, by considering *pitch-classes* $[p] = \{\, p + 12k : k \in \mathbb{Z} \,\} \cong \mathbb{R}/12\mathbb{Z}$. See Figure 6 for a representation of both the pitch and pitch-class spaces.

4.1. *A 3-dimensional, variable-geometry Tonnetz*

Originally, the *Tonnetz* was described as a graph representing the harmonic relationships among pitches, see for instance in Euler (1774). Later, it has been largely redefined, and generalized to several formalisms (Cohn 1997; Žabka 2009; Douthett and Steinbach 1998). We will consider the simplicial complex interpretation of the *Tonnetz* originally given in Bigo et al. (2013). In this setting, the *Tonnetz* is modeled as an infinite planar simplicial complex, whose 0-simplices (i.e. vertices) are labeled with pitch-classes, 1-simplices (i.e. edges) connect vertices whose labels form either perfect fifth, major, and minor third intervals. The labels associated with the vertices

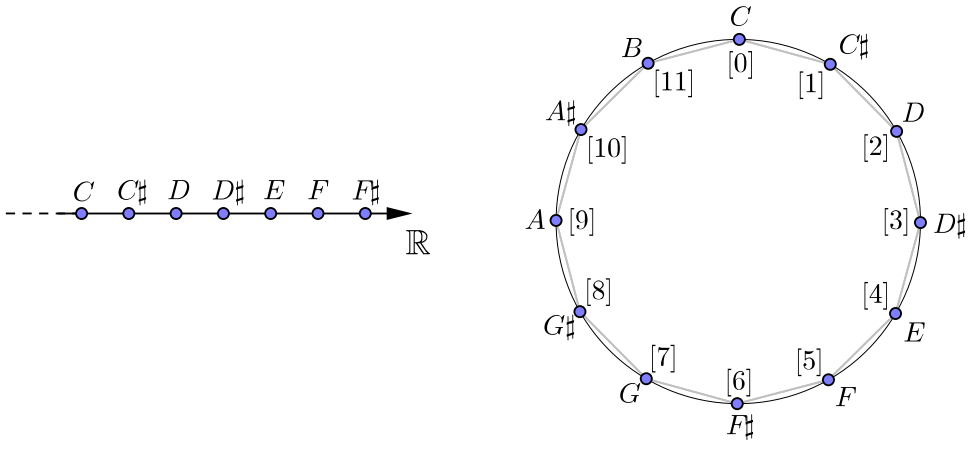


Figure 6. Fundamental music representation spaces. On the left pitches are visualized as points on the real line. The pitch-class space is visualized on the right.

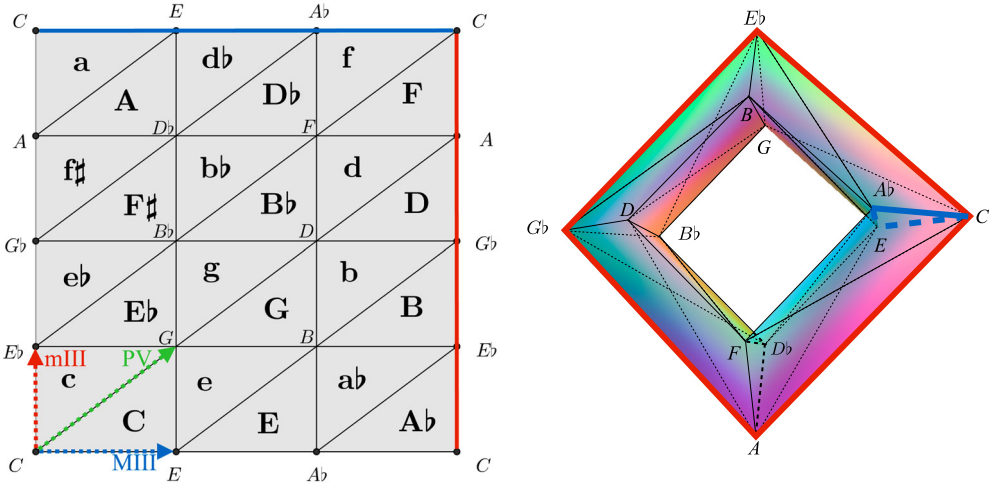


Figure 7. On the left, a finite subcomplex of the planar *Tonnetz T*. On the right, the *Tonnetz* torus $\mathbb{T}$.

of 2-simplices (i.e. triangles) correspond to either major or minor triads. A finite subcomplex of the *Tonnetz T* is depicted in Figure 7(left). We observe that the labels on its vertices are periodic with respect to the transposition of both minor and major third. This allows one to generate the finite toroidal representation $\mathbb{T}$ displayed in Figure 7(right).

It is possible to analyze and classify music by considering the subcomplexes of T generated by a sequence of pitch-classes (Bigo et al. 2013). However, this approach does not allow us to discriminate musical styles in a geometric or topological sense. Indeed, as the example in Figure 8 shows, two perceptively distinct sonorities can be represented by isomorphic subcomplexes.

Imagine the planar Tonnetz as embedded in $\mathbb{R}^3$. Then, the information concerning the harmonic relationships and the *temporal* hierarchy (durations) of notes in a musical phrase can be expressed by displacing the vertices of the *Tonnetz* along the axis normal to its surface. Moreover, music is often organised in bars. For instance, modulations generally occur every four or eight bars in a jazz context, and oftentimes melodic lines are arranged in a question and answer

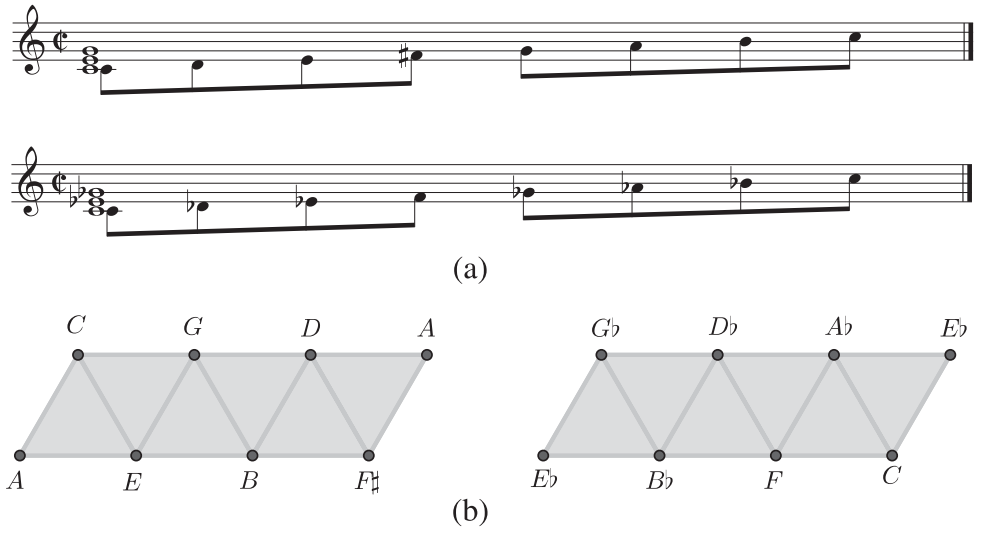


Figure 8. Two different modes represented by isomorphic subcomplexes of the *Tonnetz*. The subcomplexes are built by considering the simplices whose vertices and edges form a 2-clique community (triangles sharing an edge), such that vertices are labelled with all the pitches belonging to the chosen modal scale. (a) The lydian mode. (b) The locrian mode. (c) Ionian subcomplex and (d) Locrian subcomplex.

paradigm consisting of cycles of two or four bars in pop music. The idea is to take into account this natural segmentation to create a windowing of a given composition.

In the following paragraphs, we will describe how to create deformed versions of the planar Tonnetz by considering the pitch-classes and duration of the notes used in a composition. Thereafter, we will show how the evolution in time of a composition can be taken into account, generating a time-varying Tonnetz.

4.1.1. *Deformation of the Tonnetz*

We start by recalling that in our notation the planar Tonnetz is denoted by T, while $\mathbb{T}$ is the symbol used for its toroidal representation. Let V be the set of vertices of T. Consider the simplest case of a finite collection of notes $P = \{\mathfrak{n}_1, \ldots, \mathfrak{n}_m\} = \{(p_1, d_1), \ldots, (p_m, d_m)\}$, where p and d are the pitch and duration of each note. Assume that $\{\mathfrak{n}_{i_1}, \ldots, \mathfrak{n}_{i_k}\}$, $k \in \mathbb{N}, k \leq m$, is the subset of notes in P whose pitch belongs to the same pitch-class. Then, for every pitch-class $[p]$, we define a function

$$h : V_{[p]} \subset \mathbb{R}^3 \to \mathbb{R}^3$$
$$(x_v, y_v, 0) \mapsto (x_v, y_v, d_{v_i}),$$

where $V_{[p]} \subset V$ is the subset of vertices of T labeled with $[p]$, $d_{v_i} = \sum_{j=1}^{m} d_{l_j}^i$. The function h sums the durations of notes whose pitch belong to the same pitch-class. This can be visualised as a deformation $\overline{T}$ of the planar Tonnetz as depicted in Figure 9, panel (a). More importantly, by linear extension, it defines an ordering on the simplices of T, that allows us to induce a filtration on the finite simplicial complex $\mathbb{T}$. The filtration on $\mathbb{T}$ is consequently finite, i.e. with a finite number of homological critical values. Always in Figure 9 the configurations of a maximum, a minimum and a saddle on the geometrical realization of a portion of the deformed Tonnetz are depicted. Observe that, in our case a maximum or minimum can be a whole subcomplex of connected pitch-classes, whose vertices share the same height. A saddle can be created by playing

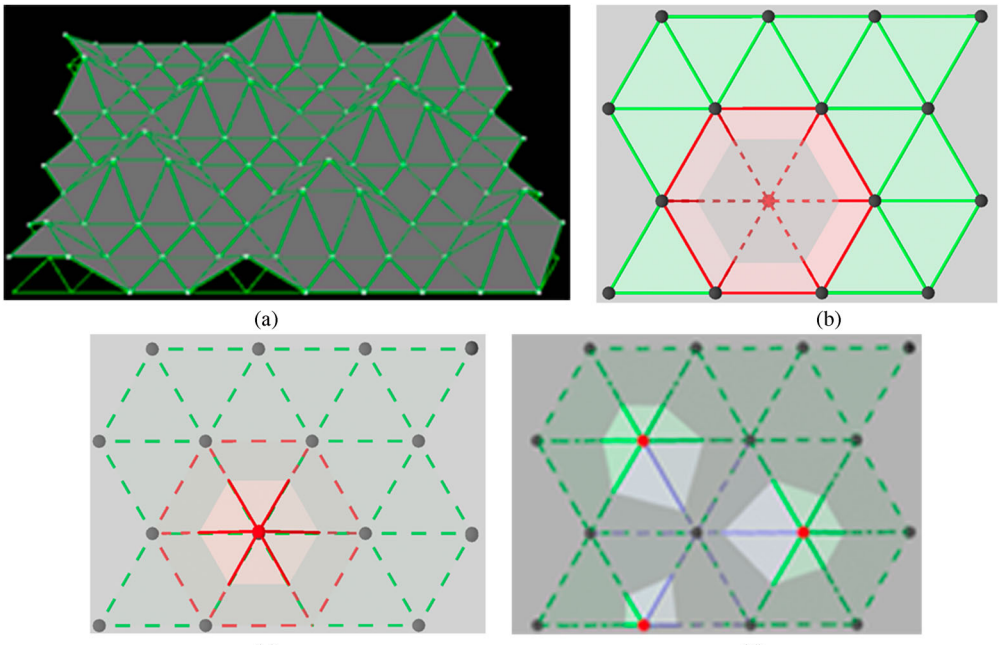


Figure 9. (a) The *Tonnetz* deformed with a major triad that appears as a 2-simplex corresponding to a maximum of the height function. Panels (b–d) represent critical points of the height function defined the deformed Tonnetz and used to induce a filtration on its toroidal counterpart $\mathbb{T}$. (b) Minimum. (c) Maximum and (d) Saddle.

a chromatic cluster of three notes (Cohn 1997). The musical interpretation of these configurations and how they shape the persistence diagrams computed on the filtration induced on $\mathbb{T}$ will be discussed in the following section.

4.1.2. *The time-varying Tonnetz*

Let $Q = \{Q_0, \ldots, Q_s\}$ be a composition consisting of $s + 1$ bars and $w \in \mathbb{N}$ be the number of bars per window we want to consider. Let us denote each window by $P_i = Q_i \cup Q_{i+1} \cup \cdots \cup Q_{i+w-1}$. Then, $Q = \cup P_i$. We can now apply the procedure described in the previous paragraph to every P_i, obtaining a time series whose observations are a deformed configuration of the *Tonnetz*. For the sake of intuition, a 3-dimensional interactive animation showing how the *Tonnetz* is deformed by a musical phrase is available at this link.[1]

The collection of deformed *Tonnetze* $\{\overline{T}_i\}$ associated with a composition will be the core of the following topological analysis. The assumption is that, given a composition, a meaningful and sufficiently refined windowing allows one to generate an accurate fingerprint of its compositional style.

5. Musical interpretation and applications

The geometry and topology of a deformed *Tonnetz* $\overline{T}$ can be considered as musical descriptors (Casey et al. 2008), and analyzed with persistent homology by defining the height function f

[1] https://mgbergomi.github.io/html/def_ton_and_sounds/examples/deformed_tonnetz_int_sound_pers.html

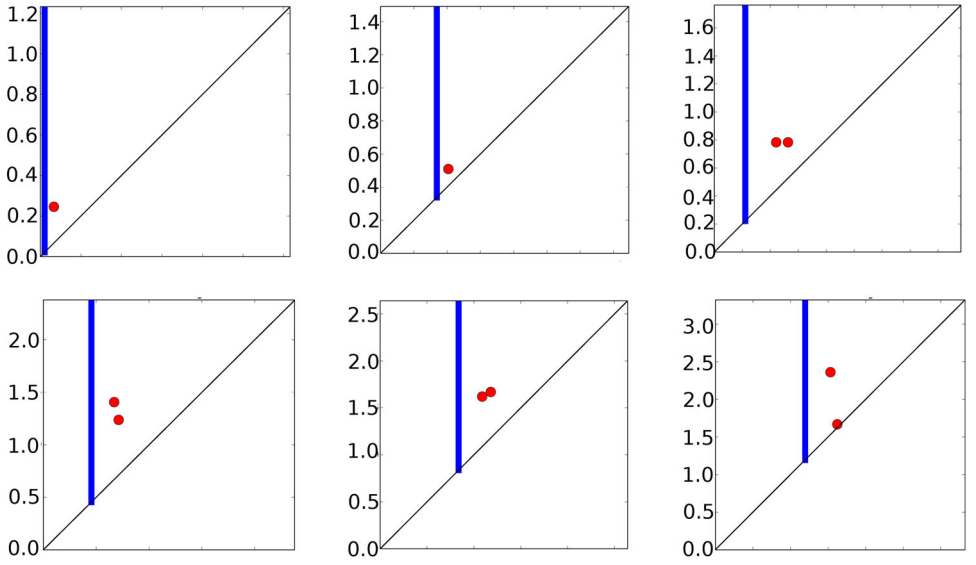


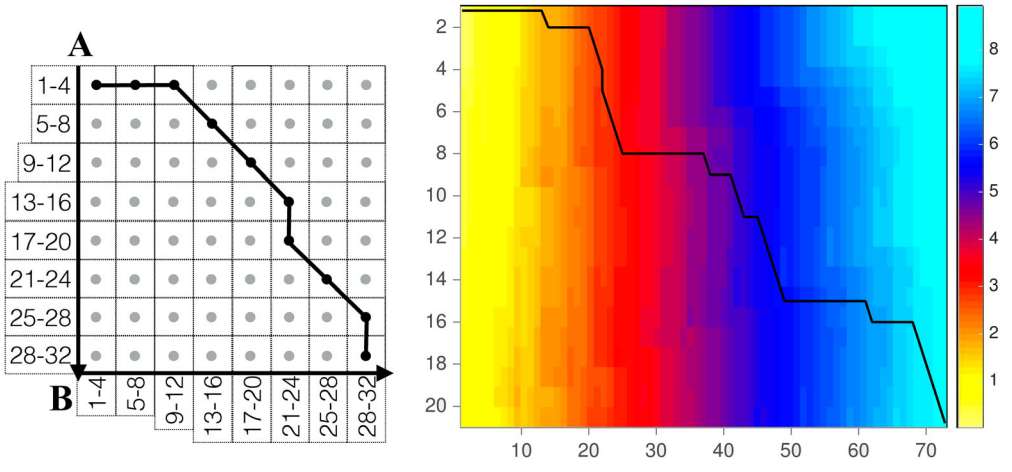
Figure 10. The first six observations of the 0-persistence time series. *Klavierstück I* – Schönberg. Persistence snapshots are taken every 8 bars.

on $\overline{T}$ to induce a lower level set filtration on the torus $\mathbb{T}$. The usage of the height function guarantees the invariance of the persistence diagrams modulo musical transposition, see Bergomi, Baratè, and Di Fabio (2016), where persistent homology has been used to fingerprint compositions as static shapes. However, this static representation does not take into account the time-dependent nature of music. We believe that persistence time series are a more natural and finer tool for the analysis of music. Indeed, this approach makes it possible to detect local relevant phenomena encoded in the progressive geometric update of the deformed *Tonnetz*. In the following applications, we will consider time series whose ith observation consists of the 0-persistence diagram associated with the space $\overline{T}_i$ defined in the previous section.

First of all, it is necessary to provide an interpretation of the music features represented as a persistence diagram.

In Figure 10, a sequence of six 0-persistence diagrams computed considering an 8-bar windowing of Schönberg's *Klavierstück I* is depicted. Consider the first diagram of the first row of the figure: a cornerline reveals the connected nature of $\mathbb{T}$ and represents the absolute minimum of the height function. This minimum corresponds to the subcomplex of the *Tonnetz* that is less used in the composition. The cornerpoint highlights the presence of a second minimum of the height function associated with a subcomplex of $\mathbb{T}$ which is disconnected from the first one. In musical terms, the presence of these two connected components can be associated with the atonal nature of the piece. The lifespan of the cornerpoint measures the relevance of this stylistic feature. The remainder of the observations describe the changes in terms of death and birth levels of these connected components. Moreover, the growth of the birth levels of the points of the whole multiset grabs the homogeneous gain of *height* of the entire simplicial complex, in time. This means that the whole chromatic scale is uniformly used in the composition, both in terms of pitches and duration of the notes. The low relative distance between cornerpoints represents the evenness, in terms of usage, of dissonant intervals during the composition: we recall how a saddle can be formed by playing three notes at a half-tone distance.

We are now ready to align two k-persistence time series, compute their dissimilarity and their optimal warping path.



(a) (b)

Figure 11. (Left) Dynamic time warping between persistence time series associated with two compositions A and B. Observations are labelled according to a 4-bars windowing. (Right) Optimal warping path between two versions of *Caravan*. The positions of the gaps correspond to the solo parts of the longer version (frames 25–50 and 51–65, respectively). (a) Optimal warping path and (b) Two versions of Caravan.

5.1. *Optimal warping path*

The first step in the calculation of an optimal warping path, given two kth persistence time series, is the computation of the pairwise bottleneck distance between their observations. The low dimensionality and simple structure of $\mathbb{T}$ assure that this normally computationally hard task can be performed in a reasonable amount of time.

In musical terms, an optimal warping path returns the comparable regions of the two compositions, represented by similar persistence diagrams with respect to the bottleneck distance. In other words, time regions of the compositions that share a similar use of the entire set of pitch-classes (cornerline), or disconnected intervals on the Tonnetz both in terms of relevance (distance from the diagonal of the cornerpoints), and in a balanced or unbalanced way (relative distance and multiplicity of the cornerpoints), are aligned in an optimal warping path. In Figure 11(a), two persistence time series associated with the compositions A and B are represented by piecewise line segments and their observations are labelled according to a 4-bars windowing. The thick line represent an optimal warping path. The first twelve bars of A are associated with the first four bars of B, in the figure an optimal warping path connects the entries $(1, 1)$ and $(1, 3)$ of the accumulated cost matrix. Symmetrically, the two last observations of A are associated with the last one of B. In this region, the warping path is represented by a vertical line segment. Diagonal line segments highlight regions of the two compositions that can be *aligned* with a reasonable cost with respect to the bottleneck distance, and hence, share similar topological features.

In the following applications, we use DTW to compute the dissimilarity between 0-persistence time series associated with three datasets composed by classical, pop, and jazz compositions, respectively. In Figure 12, the dissimilarity scores computed by aligning the compositions belonging to the datasets are depicted.

Classical dataset: Observe the first row of the figure. The pieces of the dataset are listed in Table 1. Proceed by reading the matrix from top to bottom. Both Schönberg's compositions (we will denote them by DK11-1 and DK11-2) gave high dissimilarity score when aligned with the tonal pieces. The first row of the matrix represents the dissimilarity scores computed by comparing DK11-2 with the other compositions of the dataset. The two minimal scores we retrieved

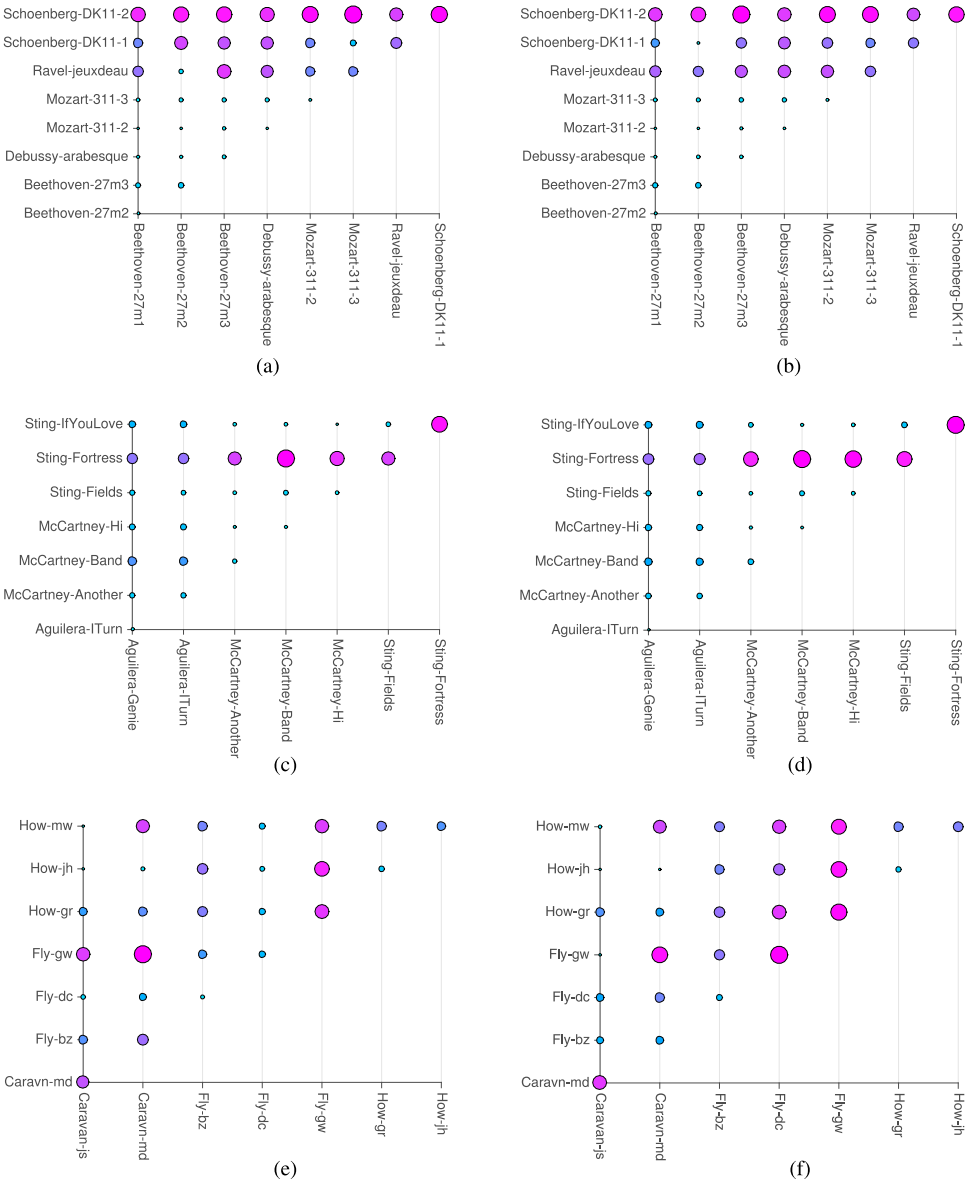


Figure 12. Alignment score of 0-persistence time series for different datasets and variable windowing. The size of the circles associated with each pair of pieces depends on their alignment score. (a) Classic Music (8 bars windowing). (b) Classic Music (4 bars windowing). (c) Pop Music (4 bars windowing). (d) Pop Music (2 bars windowing). (e) Jazz Music (8 bars windowing) and (f) Jazz Music (4 bars windowing).

are obtained by comparing DK11-2 with the compositions by Debussy and Ravel. In general, when compared to the rest of the dataset DK11-1 obtains smaller dissimilarity scores, however, they are sufficient to segregate it from the tonal pieces. The corresponding results, depicted in the distance matrix on the left, do not differ greatly from the one we just discussed. However the tonal traces left in DK11-1 are highlighted by the finer windowing we considered. The same consideration holds for the scores realized by *Jeux d'Eau*. The surprisingly low score generated by its alignment with the second movement of Beethoven's sonata changes by considering a 4-bars windowing. In this case the composition by Ravel is segregated from the others, while

Table 1. Summary of the compositions of the classical music dataset.

Composition	Movements	Author
Sonata n. 27	1, 2, 3	Beethoven
Arabesque		Debussy
Sonata n. 8	1, 2	Mozart
Jeux d'Eau		Ravel
Klavierstück	I, II	Schönberg

Table 2. Jazz dataset. Notation: B.B. arr. stands for big band arrangement, chrom. for chromatic and Man. guit. for manouche guitar.

Label	Ensemble	Style
Caravan-js	4 Gtrs, Org, Kora, Bgtr	B.B. arr., no solo
Caravan-md	Tpt, Pf, Bgtr	Rich solos and tensions
Fly-bz	Bgtr, Vib, Kora, Pf	B.B. arr., chrom. solos
Fly-dc	Flt, Tnr & Bar Sax, F Hn, Org, Gtr	B.B. arr., Man. guit.
Fly-gw	Big Band	B.B. arr.
How-gr	2 Obss, 2 Gtrs, Bgtr	Man. guit.
How-jh	Pf, Bgtr	Chrom. solo
How-mw	Tpt, Gtr, Bgtr, Str, Pf	Embellishments, chrom.

the second movement of the sonata n. 27 obtains a surprisingly low dissimilarity score when aligned with DK11-1. The tonal and pentatonic compositions are highlighted as similar in both representations.

Pop dataset: In both diagrams, the two Aguilera's pieces appear to be well separated from the others. Sting's *Fields of Gold* and *If You Love Somebody Set Them Free* turn out to be similar to the pieces by McCartney. It is not the case for *Fortress Around Your Heart* that recollects high dissimilarity scores when aligned to the other songs of the dataset. It is interesting to note how the two distance matrices are almost invariant with respect to the change of windowing.

Jazz dataset: The classification of jazz standards is a difficult task due to the improvisational nature of this genre. We considered a dataset composed of two versions of *Caravan* and three versions of *Fly Me to the Moon* and *How High the Moon*, respectively. Each interpretation is characterised by different choices in terms of ensemble and arrangements. We summarised these features in Table 2 by denoting a big band arrangement (breaks, horns fills, etc.) as *B.B. arr.*, pointing out the presence of solo parts, their main features, and particular stylistic choices.

The dissimilarity scores computed by considering the global pairwise alignment of the persistence time series associated with these compositions are depicted in the third row of Figure 12. In this example, the information retrieved by the alignment is twofold: on one hand, it stresses the mere melodic and harmonic similarity. On the other hand, it retrieves common stylistic choices. In both distance matrices, the scores associated with the same compositions are reasonably low, highlighting their similarity in the case of *Fly Me to the Moon* and *How High the Moon*. An exception is represented by the two versions of *Caravan*. The presence of rich solos in *Caravan-md* distinguishes it neatly from the other interpretation of the standard. Note how an optimal warping path between these two pieces depicted in Figure 11(b) tries to align them on the themes, skipping the solo parts. Hence, the evolution in time of the persistence diagrams grasps the difference between an organised thematic flow, and a freer improvisational context. Moreover, we notice how the three versions of *Fly Me to the Moon* is well separated from the three versions of *How High the Moon* only utilising a 4-bars windowing. This feature is opposite to the one characterising the analysis of the Pop dataset.

6. Discussion and perspectives

We presented a method to compare time-varying systems by taking advantage of their geometric and topological fingerprint provided by persistent homology.

If the two spaces and the filtering functions are comparable, DTW and an optimal warping path between two kth persistence time series represent the timespans in which the properties determined by the filtering functions, and thereafter fingerprinted as persistence diagrams, are similar. Thus, it gives an encompassing view on the relative geometric behavior of the analyzed systems, by highlighting possible irregular patterns in their geometric evolution.

The effectiveness of this dynamic fingerprint can be tested in many applications, such as animals tracking (Pérez-Escudero et al. 2014; Romero-Ferrero et al. 2019), group behaviour (Munch 2013; Topaz, Ziegelmeier, and Halverson 2015), and classical problems tackled with persistence homology as covering of sensor network (Munch, Shapiro, and Harer 2012) and unfolding of proteins (Cohen-Steiner, Edelsbrunner, and Morozov 2006).

The task of automatic stylistic classification of music is currently tackled in the field of Music Information Retrieval (MIR) (Casey et al. 2008), being an innovative paradigm for the analysis and classification of music compositions. Although the data used to deform the *Tonnetz* in our application are deduced from MIDI transcriptions, thanks to the stability of persistence diagrams with respect to the bottleneck distance, it is possible to generalize the algorithm to audio files, for instance by performing a chroma analysis (Harte and Sandler 2005).

We gave a musical interpretation of the evolution in time of the persistence diagrams associated with a composition and used DTW to provide an alignment of persistence time series. Finally, we analyzed both an optimal warping path and the alignment score of collections of classical (tonal, modal and atonal) compositions, pop songs endowed with different harmonic complexity and a collection of jazz standards played by different ensembles, with different arrangements and solo parts.

In order to interpret the evolution of a system in time as a collection of observations, we defined a windowing, consisting of an even partition of the composition according to its subdivision in bars. We saw how the three datasets we considered respond differently to changes of this windowing. In musical terms, this corresponds to the pace at which musical concepts evolve in the composition. The computation of an optimal time *granularity* necessary to describe the evolution of compositions belonging to different genres or artists could be used as a new music descriptor.

The dynamic *Tonnetz* has full memory of the composition: the heights of its vertices increases monotonically. This feature does not reflect our perception of music, since we cannot remember every note of a whole composition. The definition of a gravity function in opposition to the one generating the deformation of the vertices can be used to endow this space with a type of short-term memory. The same argument can be applied to the study of dynamical consonance-based deformed *Tonnetze*. In addition, we can define a variable gravitational field (or equip the vertices with variable masses) in order to diminish the effect of the gravity in correspondence of vertices representing relevant elements of a musical phrase (for example, its higher, lower, first, ending, and syncopated notes), we refer interested readers to Perricone (2000).

Given the simple structure of the persistence diagrams derived from the *Tonnetz* and the possibility to provide their musical interpretation in this framework, persistence time series could be replaced by continuous vineyards, in order to study the alignment between vineyards by considering the minimal homotopy leading from one to the other.

Acknowledgments

The authors acknowledge the precious comments of Massimo Ferri on an early version of this paper. We thank the reviewers for their careful reading of the manuscript, and their valuable and insightful comments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Mattia G. Bergomi http://orcid.org/0000-0002-3732-3108

References

Bakshi, B., and G. Stephanopoulos. 1994. "Representation of Process Trends-IV. Induction of Real-time Patterns From Operating Data for Diagnosis and Supervisory Control." *Computers & Chemical Engineering* 18 (4): 303–332.

Bergomi, M. G. 2015. *Dynamical and Topological Tools for (Modern) Music Analysis*. Paris: Université Pierre et Marie Curie.

Bergomi, M. G., A. Baratè, and B. Di Fabio. 2016. "Towards a Topological Fingerprint of Music." In *International Workshop on Computational Topology in Image Context*, Marseille, France, 88–100.

Berndt, D. J., and J. Clifford. 1994. "Using Dynamic Time Warping to Find Patterns in Time Series." In *KDD workshop*, 359–370. Vol. 16.

Bigo, L., M. Andreatta, J. L. Giavitto, O. Michel, and A. Spicher. 2013. "Computation and Visualization of Musical Structures in Chord-Based Simplicial Complexes." In *Mathematics and Computation in Music*, 38–51. Berlin: Springer.

Burns, E. M., and W. D. Ward. 1999. "Intervals, Scales, and Tuning." *The Psychology of Music* 2: 215–264.

Casey, M., R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. 2008. "Content-based Music Information Retrieval: Current Directions and Future Challenges." *Proceedings of the IEEE* 96 (4): 668–696.

Cheveigné, de, A. 2005. "Pitch Perception Models." In *Pitch: Neural Coding and Perception*, edited by C. J. Plack, A. J. Oxenham, A. N. Popper, and R. Fay. Springer, New York, NY.

Cohen-Steiner, D., H. Edelsbrunner, and J. Harer. 2007. "Stability of Persistence Diagrams." *Discrete & Computational Geometry* 37: 103–120.

Cohen-Steiner, D., H. Edelsbrunner, and D. Morozov. 2006. "Vines and Vineyards by Updating Persistence in Linear Time." In *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry*, 119–126. Sedona Arizona, USA. https://dl.acm.org/doi/proceedings/10.1145/1137856

Cohn, R. 1997. "Neo-Riemannian Operations, Parsimonious Trichords, and Their Tonnetz Representations." *Journal of Music Theory* 41 (1): 1–66.

d'Amico, M., P. Frosini, and C. Landi. 2006. "Using Matching Distance in Size Theory: A Survey." *International Journal of Imaging Systems and Technology* 16: 154–161.

d'Amico, M., P. Frosini, and C. Landi. 2010. "Natural Pseudo-distance and Optimal Matching Between Reduced Size Functions." *Acta Applicandae Mathematicae* 109 (2): 527–554.

Douthett, J., and P. Steinbach. 1998. "Parsimonious Graphs: A Study in Parsimony, Contextual Transformations, and Modes of Limited Transposition." *Journal of Music Theory* 42 (2): 241–263.

Edelsbrunner, H., and J. Harer. 2009. *Computational Topology: An Introduction*. American Mathematical Society.

Esling, P., and C. Agon. 2012. "Time-series Data Mining." *ACM Computing Surveys (CSUR)* 45 (1): 12.

Euler, L. 1774. "De Harmoniae Veris Principiis Per Speculum Musicum Repraesentatis." *Novi commentarii Acad. Scient. Petropolitanae* 18, 330–353, repr. in Leonhardi Euleri, Opera Omnia, Series III, T.1, Teubner, Liepzig, Berlin 1926.

Ferri, M. 2017. "Persistent Topology for Natural Data Analysis – A Survey." In *Towards Integrative Machine Learning and Knowledge Extraction*, 117–133. Cham, Switzerland: Springer.

Govc, D. 2013. "On the Definition of Homological Critical Value." arXiv:1301.6817.

Harte, C., and M. Sandler. 2005. "Automatic Chord Identifcation Using a Quantised Chromagram." In *Audio Engineering Society Convention*. Audio Engineering Society, New York, Vol. 118.

Hatcher, A. 2002. *Algebraic Topology*. Cambridge: Cambridge University Press.

Keogh, E., S. Chu, D. Hart, and M. Pazzani. 2004. "Segmenting Time Series: A Survey and Novel Approach." *Data Mining in Time Series Databases* 57: 1–22.

Keogh, E., and S. Kasetty. 2003. "On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration." *Data Mining and Knowledge Discovery* 7 (4): 349–371.

Liao, T. W. 2005. "Clustering of Time Series Data – a Survey." *Pattern Recognition* 38 (11): 1857–1874.

Lines, J., and A. Bagnall. 2015. “Time Series Classification with Ensembles of Elastic Distance Measures.” *Data Mining and Knowledge Discovery* 29 (3): 565–592.

Munch, E. 2013. *Applications of Persistent Homology to Time Varying Systems*. Durham, North Carolina: Duke University.

Munch, E., M. Shapiro, and J. Harer. 2012. “Failure Filtrations for Fenced Sensor Networks.” *The International Journal of Robotics Research* 31 (9): 1044–1056.

Munkres, J. R. 1984. *Elements of Algebraic Topology*. Vol. 2. Cambridge Massachusetts: Addison-Wesley Reading.

Perricone, J. 2000. *Melody in Songwriting: Tools and Techniques for Writing Hit Songs*. Boston: Hal Leonard Corporation.

Pérez-Escudero, A., J. Vicente-Page, R. C. Hinz, S. Arganda, and G. G. Polaviejade. 2014. “idTracker: Tracking Individuals in a Group by Automatic Identification of Unmarked Animals.” *Nature Methods* 11 (7): 743–748.

Poincaré, H. 1895. *Analysis Situs*. Paris, France: Gauthier-Villars.

Romero-Ferrero, F., M. G. Bergomi, R. C. Hinz, F. J. Heras, and G. G. Polaviejade. 2019. “idtracker. ai: Tracking All Individuals in Small Or Large Collectives of Unmarked Animals.” *Nature Methods* 16 (2): 179.

Senin, P. 2008. *Dynamic Time Warping Algorithm Review*, 40. Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA 855.

Topaz, C. M., L. Ziegelmeier, and T. Halverson. 2015. “Topological Data Analysis of Biological Aggregation Models.” *PloS One* 10 (5): e0126383.

Turner, K., Y. Mileyko, S. Mukherjee, and J. Harer. 2014. “Fréchet Means for Distributions of Persistence Diagrams.” *Discrete & Computational Geometry* 52 (1): 44–70.

Žabka, M. 2009. “Generalized Tonnetz and Well-Formed GTS: A Scale Theory Inspired by the Neo-Riemannians”. In *International Conference on Mathematics and Computation in Music*, edited by C.A.C.C.H. Chew Elaine, 286–298. Berlin: Springer.