Towards a Topological Fingerprint of Music

3 authors:

Some of the authors of this publication are also working on these related projects:

Project   Stable comparison of Reeb graphs of piecewise linear models in the financing FAR 2014 View project

Towards a topological fingerprint of music

Mattia G. Bergomi[1], Adriano Baratè[2] and Barbara Di Fabio[3]

[1] Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown,
Lisbon, Portugal,
`mattia.bergomi@neuro.fchampalimaud.org`
[2] Università degli Studi di Milano, Laboratorio di Informatica Musicale
Milano, Italy
`barate@di.unimi.it`
[3] Università di Modena e Reggio Emilia, Dipartimento di Scienze e Metodi
dell'Ingegneria
Reggio Emilia, Italy
`barbara.difabio@unimore.it`

Abstract. Can music be represented as a meaningful geometric and topological object? In this paper, we propose a strategy to describe some music features as a polyhedral surface obtained by a simplicial interpretation of the *Tonnetz*. The *Tonnetz* is a graph largely used in computational musicology to describe the harmonic relationships of notes in equal tuning. In particular, we use persistent homology in order to describe the *persistent* properties of music encoded in the aforementioned model. Both the relevance and the characteristics of this approach are discussed by analyzing some paradigmatic compositional styles. Eventually, the task of automatic music style classification is addressed by computing the hierarchical clustering of the topological fingerprints associated with some collections of compositions.

Keywords: Music, classification, clustering, Tonnetz, persistent homology

1 Introduction

Generally, the core of a piece of music consists of a small collection of strong, recognizable concepts, that are grasped by the majority of the listeners [13, 17, 29]. These *core concepts* are developed during the composition by varying levels of tension over time, drawing the attention of the listener to particular moments thanks to specific choices, frustrating his/her intuition through unexpected changes, or confirming his/her expectation with, for instance, a well-known cadence leading to resolution.

As the models for the analysis of audio signals take advantage of the strategies developed for image analysis [27, 30, 22], it is possible to borrow some tools from the topological analysis of shapes and data to tackle the problem of music analysis and classification. The main aim of this paper is the introduction of a
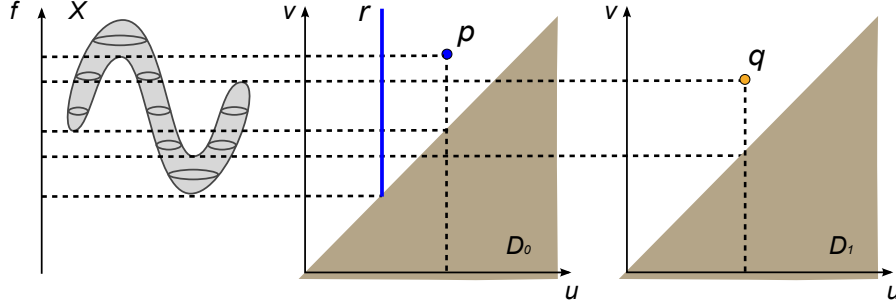
Fig. 1: Left: The height function on the topological space X. Right: The associated persistence diagrams $D_k(X, f)$, with $k = 0, 1$.

low-dimensional geometric-topological model in order to describe, albeit in an extremely simplified form, music styles.

Loosely speaking, we introduce a metric representation of music as a planar polyhedral surface, whose vertices are then translated along a third dimension in basis on a specific function. The shapes obtained via these deformations are fingerprinted by computing their *persistent homology* [15]. Afterwards, the musical meaning of this topological representation of music is discussed and applied to automatic style classification on three different datasets.

2   Background on persistent homology

In computational topology [14], persistent homology is actually considered an invaluable tool to describe both geometry and topology of a certain space, not only because of the simplicity of the method, but also because all the properties are ranked by importance, allowing us to choose the level of detail at which to perform such a description [15, 18].

In more formal terms, given a topological space X, we define a continuous function $f : X \to \mathbb{R}$ to obtain a family of subspaces $X_u = f^{-1}((-\infty, u]), u \in \mathbb{R}$, nested by inclusion, i.e. a filtration of X. The map f, called therefore a *filtering function*, is chosen according to the geometrical properties of interest (e.g., height, distance from center of mass, curvature). Applying homology to the filtration, births and deaths of topological features can be algebraically detected and their lifetime measured. The scale at which a feature is significant is measured by its longevity. Formally, given $u \leq v \in \mathbb{R}$, we consider the inclusion of X_u into X_v. This inclusion induces a homomorphism of homology groups $H_k(X_u) \to H_k(X_v)$ for every $k \in \mathbb{Z}$. The image of such a homomorphism consists of the k-homology classes that live at least from u to v along the filtration, and is called the kth *persistent homology group* of the pair (X, f) at (u, v). When, for every $(u, v), u \leq v$, the kth persistent homology groups are finitely generated, we can compactly describe them using the so-called *persistence diagrams*. A persistence diagram $D_k(X, f)$ is the subset of $\{(u, v) \in \mathbb{R}^2 : u < v\}$

consisting of points (called *proper points*) and vertical lines (called *points at infinity*) encoding the levels of f at which the birth or the death of homological classes occur, union all the points belonging to the diagonal $u = v$. In particular, if there exists at least one k-homology class that is born at the level $\bar{u}$ and is dead at the level $\bar{v}$ along the filtration induced by f, then $p = (\bar{u}, \bar{v})$ is a proper point of $D_k(X, f)$; if there exists at least one k-homology class that is born at the level $\bar{u}$ and never dies along the filtration induced by f, then $p = (\bar{u}, +\infty)$ is a point at infinity of $D_k(X, f)$. A point at infinity is usually represented as the vertical line $u = \bar{u}$. Both points and lines are equipped with a multiplicity that depends on the number of classes with the same lifetime [19]. An example of persistence diagrams is displayed in Fig. 1. The surface $X \subset \mathbb{R}^3$ is endowed with the height function f. The associated persistence diagrams $D_0(X, f)$ and $D_1(X, f)$ are displayed on the right. $D_0(X, f)$ consists of one point at infinity r, whose abscissa u detects the absolute minimum of f, and one proper point p, whose abscissa and ordinate detect, respectively, the level at which the new connected component appears and merges with the existing one. $D_1(X, f)$ consists of one proper point, whose abscissa and ordinate detect, respectively, the level at which a new tunnel is created and disappears along the filtration.

One of the main reasons behind the usage of persistence diagrams in applications consists in the possibility of estimating the degree of dissimilarity of two spaces with respect to a certain geometrical property through an appropriate comparison of these shape descriptors. Because of its properties of optimality [10] and stability [7], the most used instrument to compare persistence diagrams is given by the so called *bottleneck distance* (a.k.a. *matching distance*) [9].

Definition 1. *The bottleneck distance between two persistence diagrams D and D' is defined as*

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

where σ varies among all the bijections between D and D' and

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

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

3 Musical setting

In order to safely introduce the main model presented in this paper, we start by defining some basic musical objects.

We model a *note in equal tuning* $\mathfrak{n}$ as a pair $(p, d) \in \mathbb{R}^2$, where p is called the *pitch* of the note, and d is its *duration* in seconds. In particular, if ν denotes the fundamental frequency of $\mathfrak{n}$, the pitch $p(\nu)$ is defined as $p(\nu) = 69 + 12 \log_2 \left(\frac{\nu}{440} \right)$, where $440 Hz$ is the fundamental frequency of the note A_4 (the *la* of the fourth octave of the piano). For further details on pitches, see, e.g., [11].
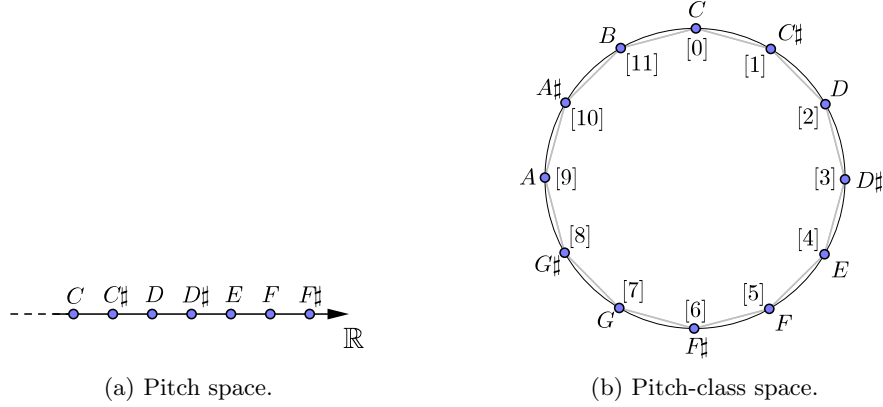
(a) Pitch space.

(b) Pitch-class space.

Fig. 2: Fundamental music representation spaces.

On a perceptual level, two notes an octave apart are really similar [4], thus, it is common to identify pitches modulo octave, by considering *pitch classes* $[p] = \{\, p + 12k : k \in \mathbb{Z} \,\} \cong \mathbb{R}/12\mathbb{Z}$. A representation of both the pitch and pitch-class spaces is depicted in Fig. 2.

3.1 The simplicial *Tonnetz*

The *Tonnetz* was originally introduced in [16] as a simple 3×4 matrix representing both the acoustical and harmonic relationships among pitch classes. Later, it has been largely generalized to several formalisms, see, e.g., [8, 32, 12]. We will focus on its interpretation as a simplicial complex [2]. In this setting, the *Tonnetz* is modeled as an infinite planar simplicial complex, whose 0-simplices are labeled with pitch classes in a way that 1-simplices form either perfect fifth, major, or minor third intervals, and 2-simplices correspond to either major or minor triads. A finite subcomplex of the *Tonnetz T* is depicted in Fig. 3a. We observe that the labels on its vertices are periodic with respect to the transposition of both minor and major third. This feature allows to work with the more comfortable toroidal representation $\mathbb{T}$ displayed in Fig. 3b.

It is possible to analyze and classify music by considering the subcomplexes of T generated by a sequence of pitch classes [2]. However, this approach does not allow to discriminate musical styles in a geometric or topological sense. In fact, as the example in Fig. 4 shows, two perceptively distinct sonorities can be represented by isomorphic subcomplexes.

3.2 A deformed *Tonnetz* for music analysis

In order to capture both the *temporal* and harmonic information encoded in a musical phrase, the vertices of the *Tonnetz* shall be displaced depending both on the pitch which is played and its duration.

4

(a) Planar simplicial *Tonnetz*.
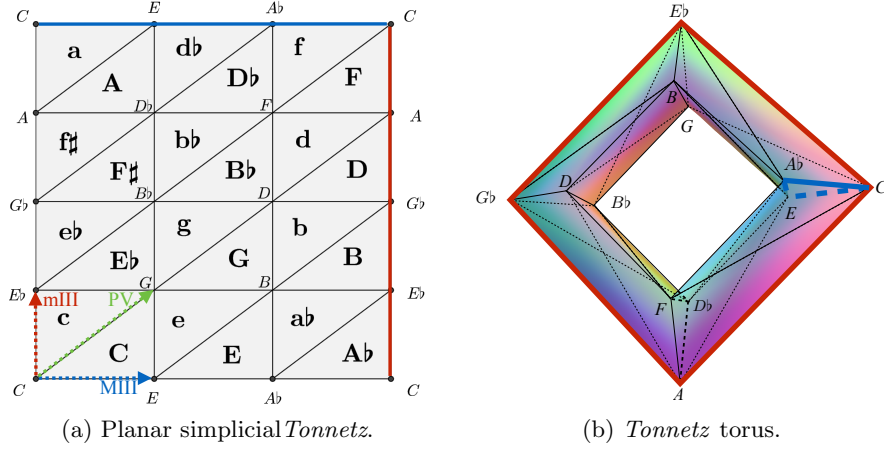
(b) *Tonnetz* torus.

Fig. 3: (a) A finite subcomplex of the *Tonnetz*. (b) The *Tonnetz* torus $\mathbb{T}$ obtained by identifying vertices in (a) equipped with the same labels.

Let V be the 0-skeleton of T and $\{\mathfrak{n}_1, \ldots, \mathfrak{n}_m\} = \{(p_1, d_1), \ldots, (p_m, d_m)\}$ a finite collection of notes of a musical phrase. Assume that $\{\mathfrak{n}_{i_1}, \ldots, \mathfrak{n}_{i_k}\}$ is the subset whose pitches $p_{i_1}, \ldots, p_{i_k}$ belong to $[p]$. We define a map that takes each vertex $v = (x_v, y_v, 0) \in \mathbb{R}^3$ labeled with $[p]$ to the point $(x_v, y_v, d_v) \in \mathbb{R}^3$, where $d_v = \sum_{j=1}^{k} d_{i_j}$, and then extend it linearly to all the simplices. The *Tonnetz* deformed under the action of this map will be denoted by $\mathcal{T}$, and will be used as the main object of our topological description of music style. An example of deformation induced by a major triad played for 8 seconds is depicted in Fig. 5, while a 3-dimensional interactive animation showing how the *Tonnetz* is deformed by a musical phrase is available at `http://nami-lab.com/tonnetz/examples/deformed_tonnetz_int_sound_pers.html`.

4 A topological fingerprint of music styles

In order to describe the deformed *Tonnetz*, we use persistent homology.

We define the height function f on $\mathcal{T}$ to induce a lower level set filtration on the torus $\mathbb{T}$. The persistence diagrams obtained with this process are *descriptors* of the style characterizing the composition represented as a shape.

0th persistence diagrams. The connectedness of $\mathbb{T}$ is retrieved by the presence of only one point at infinity. Let $u = \bar{u}$ be its equation: $\bar{u}$ is the absolute minimum of f on the deformed *Tonnetz*. If $\bar{u} \approx 0$, then there exists at least one pitch-class set that does not have a relevant role in the composition, suggesting that it is based on a stable tonal or modal choice. On the contrary, if $\bar{u} >> 0$, then all the pitch classes have been used in the composition for a relevant time. This configuration suits a more atonal or chromatic style. The presence of proper points is due to the existence of minima of the height function, that are
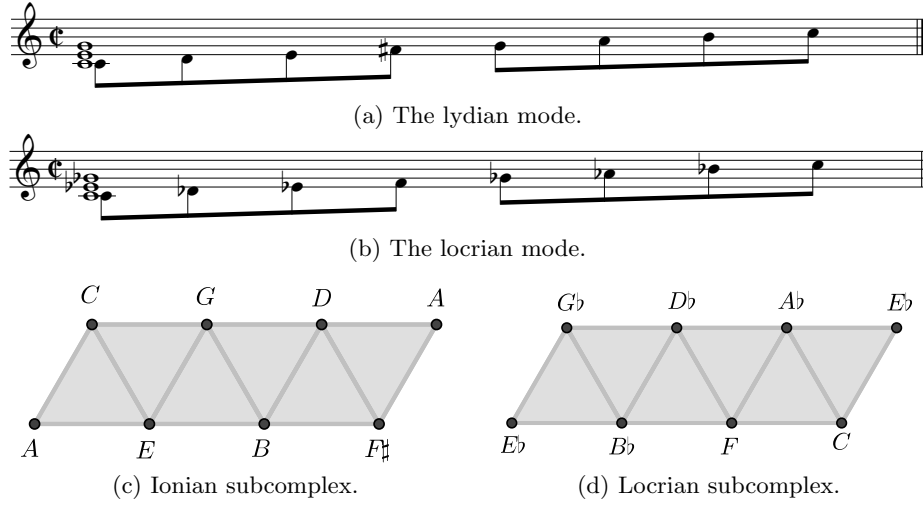
5

(a) The lydian mode.

(b) The locrian mode.

(c) Ionian subcomplex.

(d) Locrian subcomplex.

Fig. 4: Two different modes represented by isomorphic subcomplexes of the *Tonnetz*.

(a) $\mathcal{T}$
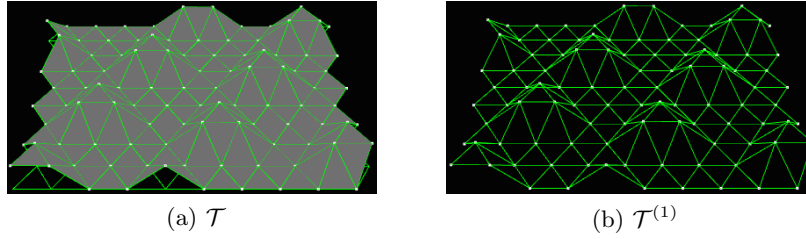

(b) $\mathcal{T}^{(1)}$

Fig. 5: The *Tonnetz* deformed with a major triad (a) and its 1-skeleton (b). The triad appears as a maximal triangle with respect to the height function.

subcomplexes of the *Tonnetz* not connected by an edge, and hence, representing a dissonant interval [26]. Furthermore, the structure of the *Tonnetz* torus allows to retrieve a maximum of three connected components. To create this particular configuration, it is necessary to play a chromatic cluster: for instance, $C, C\sharp, D$, that is not usually used in a tonal or modal context.

1st persistence diagrams. The lifespan of 1-dimensional holes traversing the filtration provides symmetrical information with respect to the 0th persistence analysis. In this case, two points at infinity detect the two generators of the 1st homology group of the torus and, if there exists, proper points detect the presence of maxima of the height function, that are subcomplexes of the *Tonnetz* not connected by an edge.

As an example, we consider the persistence diagrams associated with Debussy's *Arabesque, Jeux d'Eau* by Ravel, and *Klavierstück 1* by Schönberg, shown in Fig. 6. In the 0th persistence diagram describing *Arabesque*, there
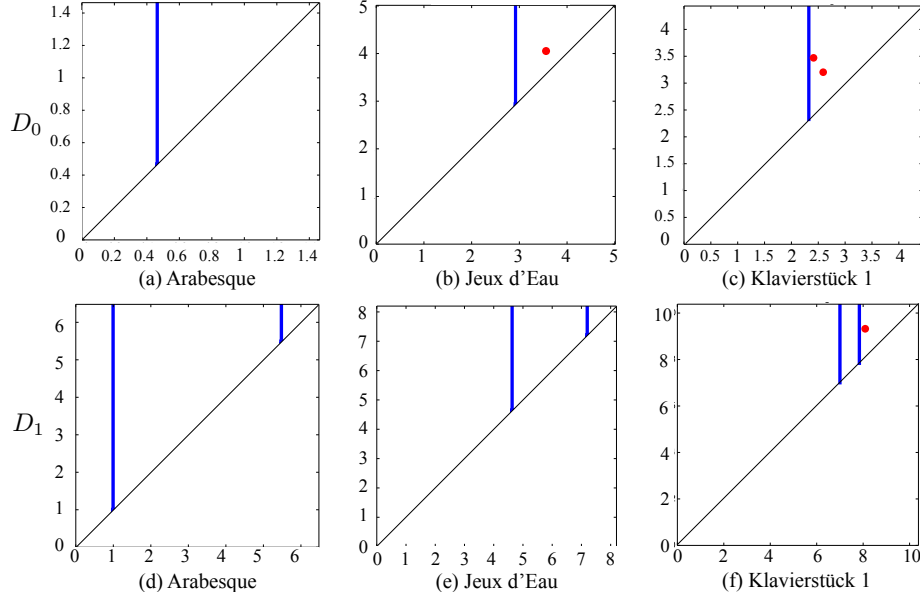
6

Fig. 6: The 0 (first row) and 1st persistence diagrams (second row) representing the topological fingerprints associated with three different compositions.

are no proper points. This is an evidence of the pentatonic and diatonic/modal inspiration of the composition [28]. We also observe that the entire chromatic scale has been used, since $\bar{u} > 0$. The abscissa of the point at infinity in the 0th persistence diagram of *Jeux d'Eau* is characterized by a high value, thus the entire set of pitch classes has been largely used in the composition. Moreover, the presence of a proper point highlights the *ante-litteram* use of the *Petrushka chord*, a superposition of a major triad and its tritone substitute: for instance, $G = (G, B, D) + C\sharp = (C\sharp, E\sharp, G\sharp)$. Finally, the diagram associated with the *Klavierstück 1* has two relevant proper points: this last feature points out the atonal nature of the composition.

The second row of Fig. 6 shows the 1st persistence diagrams associated with the same compositions. The tonal nature of *Arabesque* is highlighted by the a large distance between the points at infinity and the absence of proper points. The chromatic style of *Jeux d'Eau* implies the reduction of the distance between the two points at infinity. This last feature appears also in the diagram describing the *Klavierstück 1*, whose atonal tendency is stressed by a proper point representing the relevant lifespan of a third non-connected subcomplex.

4.1 Applications

In the following applications, we show how the persistence diagrams associated with a collection of compositions can effectively classify them according to their
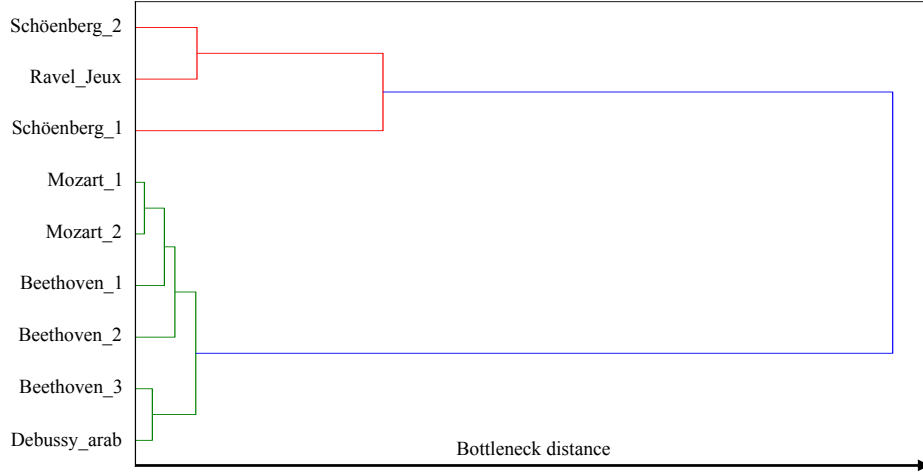
Fig. 7: Persistence-based clustering of nine classical and contemporary pieces.

style. For $k = 0, 1$, let $\mathcal{D} = \{D_1, \ldots, D_n\}$ be the set of kth persistence diagrams. Let $M = m_{ij} = d_B(D_i, D_j)$, for $1 \leqslant i, j \leqslant n$, be their distance matrix. The hierarchical clustering analysis [25] allows us to describe the configuration of the diagrams $D_i \in \mathcal{D}$ with respect to the bottleneck distance. We will represent the organization of all their possible clusters as a dendrogram [21, 23]. In this type of diagram, the abscissa of each splitting (vertical line) measures the distance between two clusters. Such distance is computed through elementary operations on the elements of M.

Tonal and Atonal Music. We consider a dataset composed by nine pieces selected among the compositions by Beethoven, Debussy, Mozart, Ravel and Schönberg available at `http://nami-lab.com/tonnetz/examples/deformed_tonnetz_int_sound_pers.html`.

The clustering computed using the 0th persistence of these pieces is depicted in Fig. 7. Data are organized in two main clusters, that segregate the two first pieces of Schönberg's *Drei Klavierstücke* and Ravel's *Jeux d'Eau*, from the ones by Mozart, Beethoven and Debussy. The association between *Klavierstück 2* and *Jeux d'Eau* mirrors the particular nature of this Schönberg's composition, that lies at the crossroad of tonal and atonal music, as it is proven by its disparate tonal interpretations [3, 31, 24]. The two movements of Mozart's *KV311* form immediately a cluster reached at an increasing distance by the two first movements of the *Sonata in C major* by Beethoven. The third movement of *Sonata in C major* is grouped with *Arabesque* because both are characterized by a generous use of the pentatonic scale.

Comparing three versions of *All the Things You Are*. The three interpretations are structured as follows:
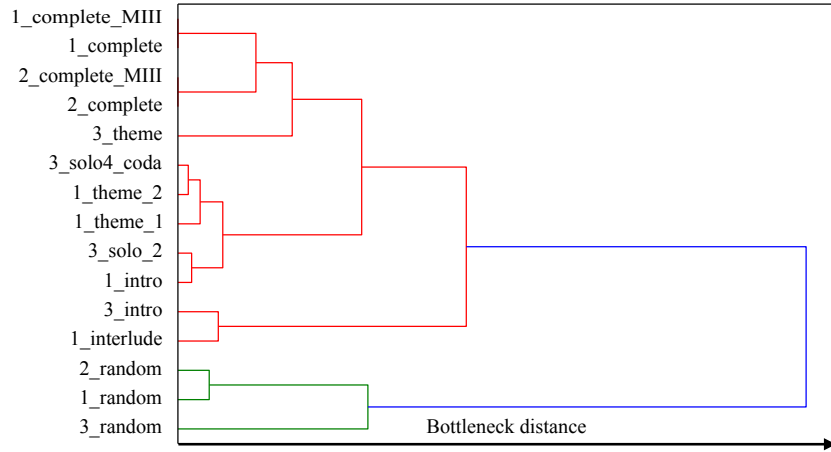
Fig. 8: Comparing three different versions of All the Things You Are.

1. Version 1 is played by a quartet in a standard way. Its main sections are a 3/4 introduction, a first exposition of the theme, and a 12 bars interlude introducing the last theme enriched by short improvisations.
2. Version 2 is performed by a piano solo, and it is characterized by a rich chromatic playing style of both hands. The main theme is executed twice.
3. Version 3 is performed by a duo (piano and bass). Its structure consists of an introduction, an exposition of the theme, and a piano improvisation.

The dataset is composed by the complete versions of the standard labeled as *i_complete* and a transposed version (*i_complete_interval*). Segments of each versions are included in the dataset and labeled as *i_segment*. In order to test the ability of the model to distinguish between a piece modulating in several tonalities, enriched with chromatic solos, and a non-structured sequence of pitches, a random version of each interpretation is also part of the dataset (*i_random*).

The resulting dendrogram is displayed in Fig. 8. We observe that the transposed versions have distance zero from the original ones, as an effect of the invariance of the filtration induced on the *Tonnetz* torus by the height function under uniform transposition. The randomized versions of the songs are well segregated. A small cluster groups the interlude of the first and the introduction of the third version, because both fragments share a very similar structure in terms of intervallic leaps and rhythm. Finally, in the top cluster, the two complete songs are linked to the fragment of the third version containing the theme. Hence, the 0th persistence homology retrieves the fragments containing the whole structure of the standard. This feature is surprising when taking into account the several modulations of the piece.

Big Pop Clustering. Fig. 9 shows a simplification of the clustering resulting by the comparison of the 1st persistence diagrams associated with 58 pop
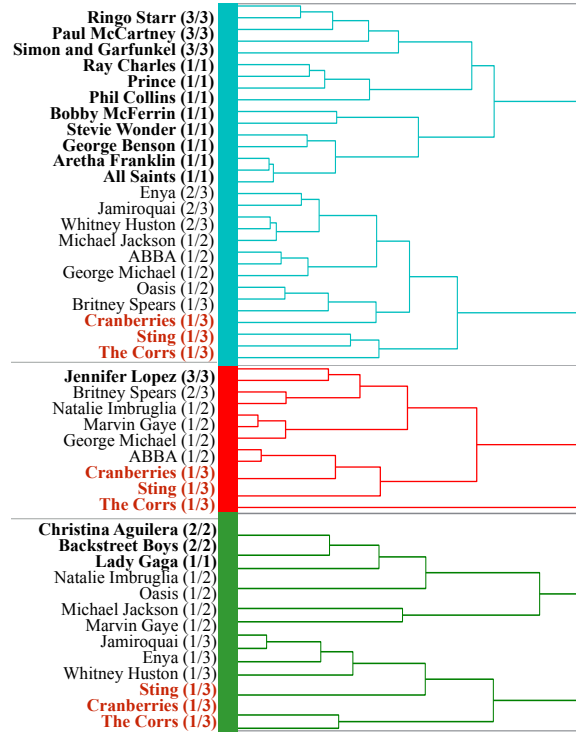
9

Fig. 9: A simplified representation of the clustering of 58 pop songs generated from their 1st persistence diagrams.

songs performed by 28 artists, spacing from Ray Charles to Lady Gaga. In order to give a simplified representation of this dendrogram, we considered only the three biggest clusters detected by the algorithm. On the left of each cluster, we listed the artists whose songs belong to that group. Names written in black bold characters indicate artists whose songs are entirely grouped in the cluster at their right, while red bold characters identify the three artists whose songs are spread among the three groups. We observe how the entire collection of songs by Ringo Starr, Paul McCartney and Simon & Garfunkel are grouped together in the blue cluster with Ray Charles, Stevie Wonder and George Benson. Moreover, the heterogeneity that characterizes Sting's compositions is mirrored by the presence of one of his songs in each cluster. The second and third clusters are less homogeneous, but promising, taking into account that so far each song is identified by a single persistence diagram.

5 Discussion and future works

We suggested a model describing music by taking into account the contribution of each pair (pitch class, duration) associated with the notes of a composition.

10

The height function has been defined on the vertices the simplicial *Tonnetz* to induce a lower level set filtration on the *Tonnetz* torus. The 0th and 1st persistence diagrams associated with different musical pieces have been interpreted in musical terms and their bottleneck distance has been used to classify them hierarchically. The possible clusterings have been represented as dendrograms, showing that 0th and 1st persistence can be used to analyze and classify music.

The analysis and classification of music we performed has been realized by considering datasets composed by MIDI files. However, the extension of this model to audio files is straightforward. Given an audio signal, the chroma analysis [20] retrieves the contribution in time of each pitch class. Using a chromagram to define the height function, it would surely be affected by the noisy data coming form the signal. However, the stability of the persistence diagrams, when compared using the bottleneck distance, assures that a small perturbation of the function inducing the filtration corresponds to small variations of the persistence diagrams.

The model itself can be extended in several ways. For instance, it is possible to augment the dimensionality the simplicial *Tonnetz*. This would result in losing its property to be easily visualizable, but it would give the possibility to encode more information. This could be done by associating with each pitch class of the *Tonnetz* a velocity, or by adding information concerning whatever pitch-class related feature. Moreover, topological persistence offers further tools to improve the strategies we suggested. A natural development is the study of the multidimensional persistent homology [5, 6] of musical spaces and their time-varying nature [1].

References

1. Bergomi, M.G.: Dynamical and Topological Tools for (Modern) Music Analysis. Ph.D. thesis, Université Pierre et Marie Curie (2015)
2. Bigo, L., Andreatta, M., Giavitto, J.L., Michel, O., Spicher, A.: Computation and visualization of musical structures in chord-based simplicial complexes. In: Mathematics and Computation in Music, pp. 38–51. Springer (2013)
3. Brinkmann, R.: Arnold Schönberg, drei Klavierstücke Op. 11: Studien zur frühen Atonalität bei Schönberg. Franz Steiner Verlag (1969)
4. Burns, E.M., Ward, W.D.: Intervals, scales, and tuning. The psychology of music 2, 215–264 (1999)
5. Carlsson, G., Zomorodian, A.: The theory of multidimensional persistence. Discrete & Computational Geometry 42(1), 71–93 (2009)
6. Cerri, A., Di Fabio, B., Ferri, M., Frosini, P., Landi, C.: Betti numbers in multidimensional persistent homology are stable functions. Mathematical Methods in the Applied Sciences 36(12), 1543–1557 (2013)
7. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. Discrete & Computational Geometry 37(1), 103–120 (2007)
8. Cohn, R.: Neo-riemannian operations, parsimonious trichords, and their" tonnetz" representations. Journal of Music Theory pp. 1–66 (1997)
9. d'Amico, M., Frosini, P., Landi, C.: Using matching distance in size theory: A survey. Int. J. Imag. Syst. Tech. 16(5), 154–161 (2006)

10. d'Amico, M., Frosini, P., Landi, C.: Natural pseudo-distance and optimal matching between reduced size functions. Acta Applicandae Mathematicae 109(2), 527–554 (2010)
11. De Cheveigne, A.: Pitch perception models. In: Pitch, pp. 169–233. Springer (2005)
12. Douthett, J., Steinbach, P.: Parsimonious graphs: A study in parsimony, contextual transformations, and modes of limited transposition. Journal of Music Theory pp. 241–263 (1998)
13. Dowling, W.J.: Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. Perception & Psychophysics 12(5), 417–421 (1972)
14. Edelsbrunner, H., Harer, J.: Computational Topology: An Introduction. American Mathematical Society (2009)
15. Edelsbrunner, H., Harer, J.: Persistent homology-a survey. Contemporary mathematics 453, 257–282 (2008)
16. Euler, L.: De harmoniae veris principiis per speculum musicum repraesentatis. Opera Omnia 3(1), 568–586 (1774)
17. Folgieri, R., Bergomi, M.G., Castellani, S.: EEG-Based Brain-Computer Interface for Emotional Involvement in Games Through Music. In: Digital Da Vinci, pp. 205–236. Springer (2014)
18. Frosini, P., Landi, C.: Size theory as a topological tool for computer vision. Pattern Recogn. and Image Anal. 9, 596–603 (1999)
19. Frosini, P., Landi, C.: Size functions and formal series. Appl. Algebra Engrg. Comm. Comput. 12(4), 327–349 (2001)
20. Harte, C., Sandler, M.: Automatic chord identifcation using a quantised chromagram. In: Audio Engineering Society Convention 118. Audio Engineering Society (2005)
21. Langfelder, P., Zhang, B., Horvath, S.: Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24(5), 719–720 (2008)
22. Li, T.L., Chan, A.B., Chun, A.: Automatic musical pattern feature extraction using convolutional neural network. In: Proc. Int. Conf. Data Mining and Applications (2010)
23. Martinez, W.L., Martinez, A., Solka, J.: Exploratory data analysis with MATLAB. CRC Press (2010)
24. Ogdon, W.: How Tonality Functions In Shoenberg Opus-11, Number-1. Journal of the Arnold Schoenberg Institute 5(2), 169–181 (1981)
25. Ott, N.: Visualization of Hierarchical Clustering: Graph Types and Software Tools. GRIN Verlag (2009)
26. Plomp, R., Levelt, W.J.: Tonal consonance and critical bandwidth. Journal of the Acoustical Society of America 38(4), 548–560 (1965)
27. Smaragdis, P., Brown, J.C.: Non-negative matrix factorization for polyphonic music transcription. In: Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on. pp. 177–180. IEEE (2003)
28. Trezise, S.: The Cambridge Companion to Debussy. Cambridge University Press (2003)
29. Tulipano, L., Bergomi, M.G.: Meaning, music and emotions: a neural activity analysis. In: NEA Science. pp. 105–108 (2015)
30. Wang, A., et al.: An Industrial Strength Audio Search Algorithm. In: ISMIR. pp. 7–13 (2003)
31. William, B.: Harmony in Radical European Music. Society of Music Theory (1984)
32. Zabka, M.: Generalized Tonnetz and Well-Formed GTS: A Scale Theory Inspired by the Neo-Riemannians? Mathematics and Computation in Music p. 286 (2009)