

Towards a Topological Fingerprint of Music

Mattia G. Bergomi¹(✉), Adriano Baratè², and Barbara Di Fabio³

¹ Champalimaud Neuroscience Programme,
Champalimaud Centre for the Unknown, Lisbon, Portugal
mattia.bergomi@neuro.fchampalimaud.org

² Laboratorio di Informatica Musicale,
Università degli Studi di Milano, Milano, Italy
barate@di.unimi.it

³ Dipartimento di Scienze e Metodi dell'Ingegneria,
Università di Modena e Reggio Emilia, 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 [22, 27, 30], 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 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 shifted 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 $f : X \rightarrow \mathbb{R}$ 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 $H_k(X_u) \rightarrow H_k(X_v)$ 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 k th *persistent homology group* of the pair (X, f) at (u, v) . When, for every (u, v) , $u \leq v$, the k th 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 and 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

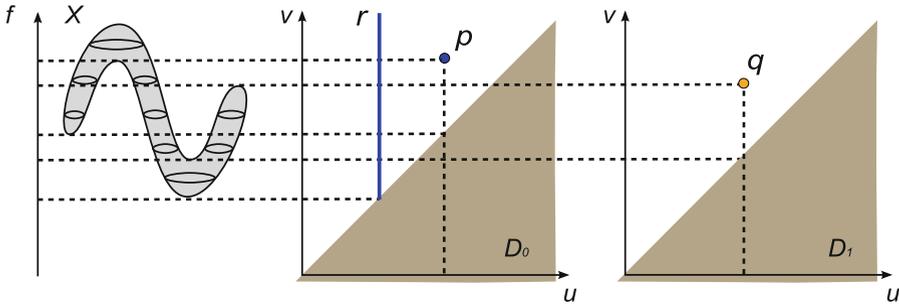


Fig. 1. Left: the height function on the topological space X . Right: the associated persistence diagrams $D_k(X, f)$, with $k = 0, 1$.

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 stability [7] and optimality [10], the most used instrument to compare persistence diagrams is given by the so called *bottleneck distance* (a.k.a. *matching distance*) [9]. By stability, we mean that small changes of the filtering functions imply only small changes of the associated persistence diagrams in terms of this distance; by stability and optimality, we mean that, among all the stable distances between persistence diagrams, the bottleneck distance is the most discriminative one.

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

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

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.

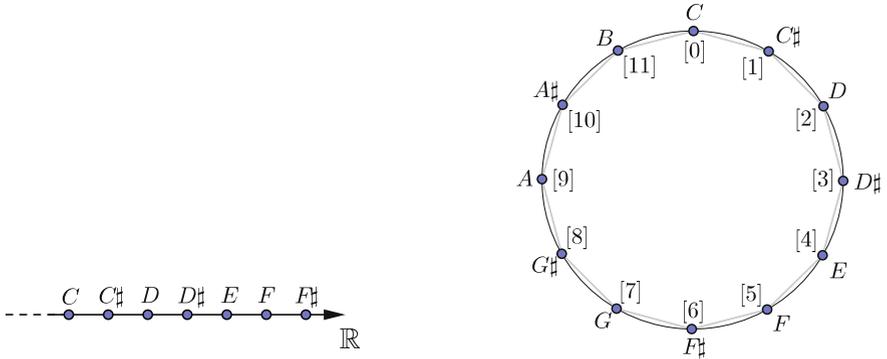


Fig. 2. Fundamental music representation spaces: the pitch space (left), and the pitch-class space (right).

We model a *note in equal tuning* \mathbf{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 \mathbf{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].

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, 12, 32]. 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 (Fig. 4a and b) can be represented by isomorphic subcomplexes (Fig. 4c and d, respectively).

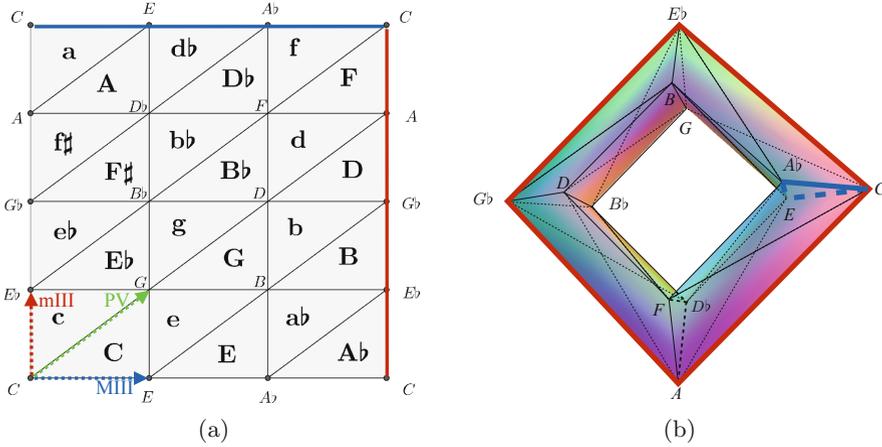


Fig. 3. (a) A finite subcomplex of the *Tonnetz*. (b) The *Tonnetz* torus \mathbb{T} obtained by identifying vertices in (a) equipped with the same labels.

3.2 A Deformed *Tonnetz* for Music Analysis

In order to capture both the *temporal* and harmonic information encoded in a musical phrase, the associated *Tonnetz* should depend on both the pitches which are played and their duration. Our idea is based on the following observation: Let us consider again the example in Fig. 4. The musical phrases in Fig. 4a and b can be easily distinguished if we endow each vertex of the associated subcomplexes with a non-negative real number that detects for how long the associated pitches have been played during the execution of the phrase. Practically, we can replace the planar subcomplexes in Fig. 4c and d by the new subcomplexes Fig. 4e and f, respectively, obtained by shifting upward each vertex of a quantity equal to this weight.

In symbols, let V be the set of vertices of T and consider a finite collection of notes of a musical phrase, $\{\mathbf{n}_1, \dots, \mathbf{n}_m\} = \{(p_1, d_1), \dots, (p_m, d_m)\}$. Assume that $\{\mathbf{n}_{i_1}, \dots, \mathbf{n}_{i_k}\}$ is the subset whose pitches p_{i_1}, \dots, 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. 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. It allows the user to play with its own keyboard to generate specific deformations or simply to analyze their evolution in real time during the execution of a certain song.

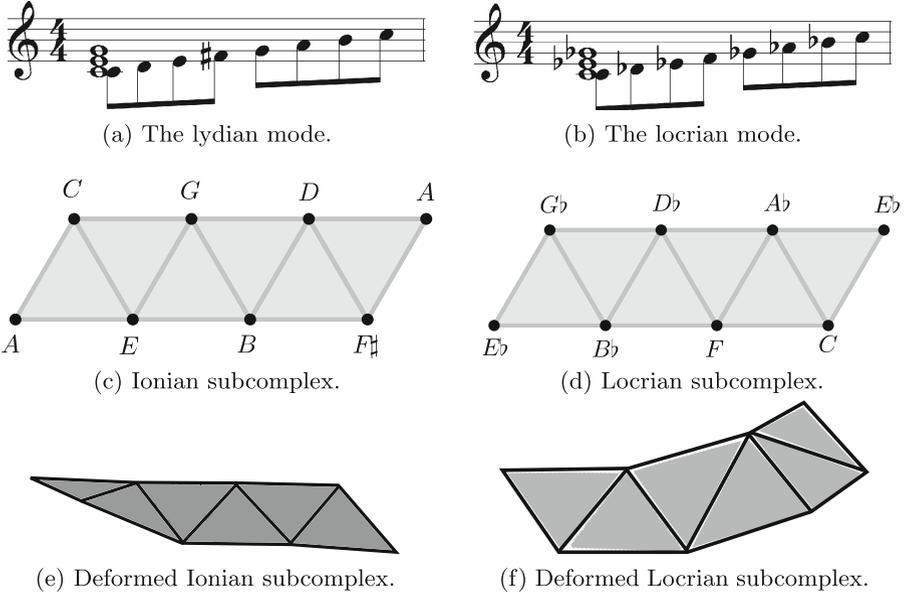


Fig. 4. Two modes (a) and (b) represented by isomorphic subcomplexes (c) and (d) of the *Tonnetz*. They can be distinguished when deforming the *Tonnetz* by taking into account the duration of the notes: (e) and (f) are deformed subcomplexes associated to the modes, observed from the same position.

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 $\bar{u} \gg 0$, 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 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

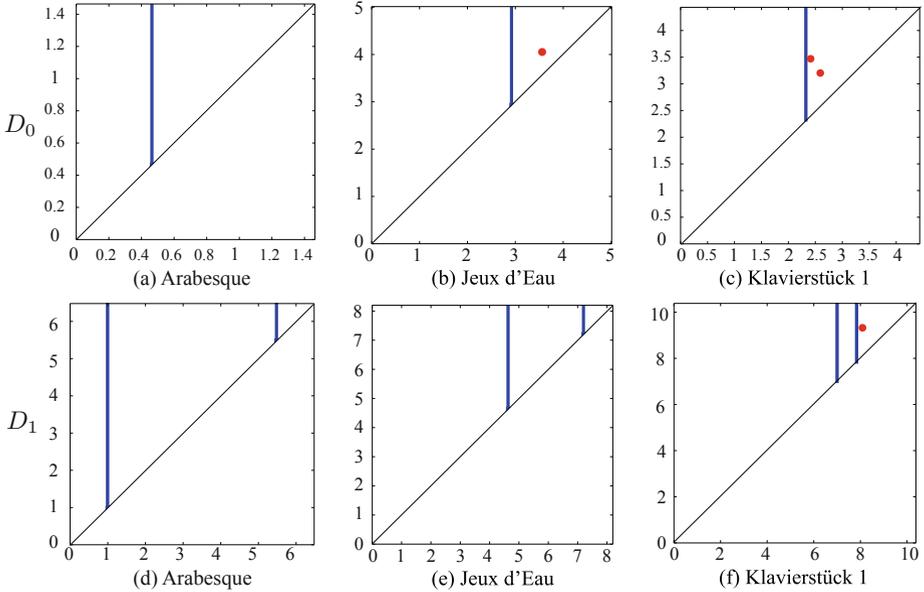


Fig. 5. The 0th (first row) and 1st persistence diagrams (second row) representing the topological fingerprints associated with three different compositions.

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. 5. In the 0th persistence diagram describing *Arabesque*, there 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. 5 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.

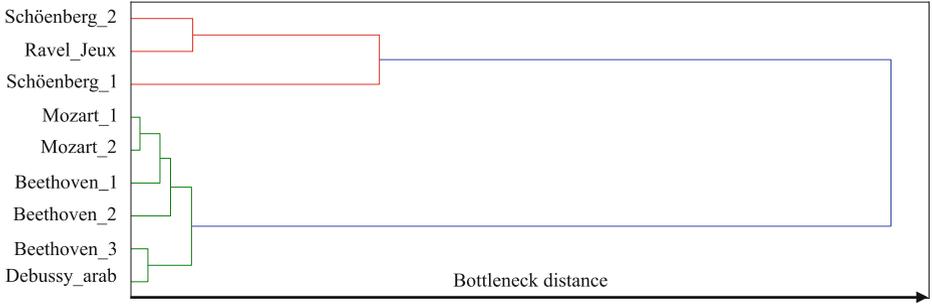


Fig. 6. Persistence-based clustering of nine classical and contemporary pieces.

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 style. For $k = 0, 1$, let $\mathcal{D} = \{D_1, \dots, D_n\}$ be the set of k th persistence diagrams. Let $M = m_{ij} = d_B(D_i, D_j)$, for $1 \leq i, j \leq n$, 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. 6. 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, 24, 31]. 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:

Big Pop Clustering. Figure 8 shows a simplification of the clustering resulting by the comparison of the 1st persistence diagrams associated with 58 pop 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.

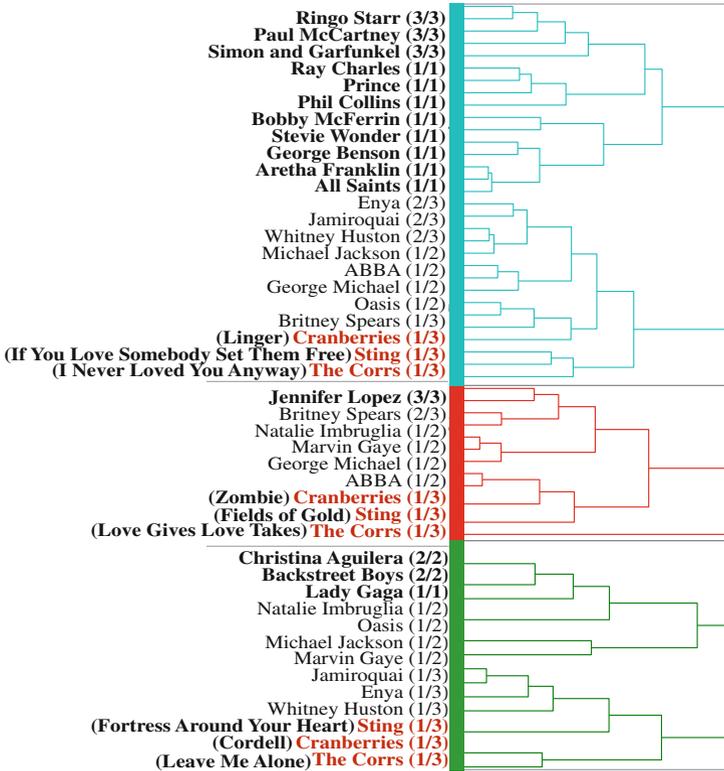


Fig. 8. A simplified representation of the clustering of 58 pop songs generated from their 1st persistence diagrams.

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. The height function has been defined on the vertices of 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 from the signal. The stability of persistence diagrams, when compared using the bottleneck distance, assures robustness against noise.

The model itself can be extended in several ways. For instance, it is possible to augment the dimensionality of the simplicial *Tonnetz*. This would result in losing the property of being 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: Yust, J., Wild, J., Burgoyne, J.A. (eds.) MCM 2013. LNCS, vol. 7937, pp. 38–51. Springer, Heidelberg (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. *Psychol. Music* **2**, 215–264 (1999)
5. Carlsson, G., Zomorodian, A.: The theory of multidimensional persistence. *Discrete Comput. Geom.* **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. *Math. Meth. Appl. Sci.* **36**(12), 1543–1557 (2013)
7. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. *Discrete Comput. Geom.* **37**(1), 103–120 (2007)

8. Cohn, R.: Neo-riemannian operations, parsimonious trichords, and their “Tonnetz” representations. *J. Music Theor.* **41**, 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: Plack, C.J., Fay, R.R., Oxenham, A.J., Popper, A.N. (eds.) *Pitch*, pp. 169–233. Springer, New York (2005)
12. Douthett, J., Steinbach, P.: Parsimonious graphs: a study in parsimony, contextual transformations, and modes of limited transposition. *J. Music Theor.* **42**, 241–263 (1998)
13. Dowling, W.J.: Recognition of melodic transformations: inversion, retrograde, and retrograde inversion. *Percept. Psychophys.* **12**(5), 417–421 (1972)
14. Edelsbrunner, H., Harer, J.: *Computational Topology: An Introduction*. American Mathematical Society, Providence (2009)
15. Edelsbrunner, H., Harer, J.: Persistent homology—a survey. *Contemp. Math.* **453**, 257–282 (2008)
16. Euler, L.: *De harmoniae veris principiis per speculum musicum repraesentatis*. *Opera Omnia* **3**(2), 568–586 (1774)
17. Folgieri, R., Bergomi, M.G., Castellani, S.: EEG-based brain-computer interface for emotional involvement in games through music. In: Lee, N. (ed.) *Digital Da Vinci. Computers in Music*, pp. 205–236. Springer, New York (2014)
18. Frosini, P., Landi, C.: Size theory as a topological tool for computer vision. *Pattern Recogn. Image Anal.* **9**, 596–603 (1999)
19. Frosini, P., Landi, C.: Size functions and formal series. *Appl. Algebra Eng. Comm. Comput.* **12**(4), 327–349 (2001)
20. Harte, C., Sandler, M.: Automatic chord identification 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: *Proceedings of International Conference on Data Mining and Applications* (2010)
23. Martinez, W.L., Martinez, A., Solka, J.: *Exploratory Data Analysis with MATLAB*. CRC Press, Boca Raton (2010)
24. Ogdon, W.: How tonality functions in Schoenberg Opus-11, Number-1. *J. Arnold Schoenberg Inst.* **5**(2), 169–181 (1981)
25. Ott, N.: *Visualization of Hierarchical Clustering: Graph Types and Software Tools*. GRIN Verlag, Munich (2009)
26. Plomp, R., Levelt, W.J.: Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* **38**(4), 548–560 (1965)
27. Smaragdis, P., Brown, J.C.: Non-negative matrix factorization for polyphonic music transcription. In: *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 177–180. IEEE (2003)
28. Trezise, S.: *The Cambridge Companion to Debussy*. Cambridge University Press, Cambridge (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. In: Society of Music Theory (1984)
32. Żabka, M.: Generalized *Tonnetz* and well-formed GTS: a scale theory inspired by the neo-riemannians. In: Chew, E., Childs, A., Chuan, C.-H. (eds.) MCM 2009. CCIS, vol. 38, pp. 286–298. Springer, Heidelberg (2009)