Topological modeling for musical structures and processes

Victoria Callet

Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

École doctorale Mathématiques, sciences de l'information et de l'ingénieur | ED 269 Université de Strasbourg

ÉCOLE DOCTORALE MATHEMATIQUES, SCIENCES DE L'INFORMATION ET DE L'INGENIEUR – ED269

Institut de Recherche Mathématiques Avancées – UMR 7501

THÈSE présentée par :

Victoria Callet

soutenue le : **15 décembre 2023**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/Spécialité : Mathématiques

Modélisation topologique des structures et processus musicaux

THÈSE dirigée par :

Pierre Guillot	Maître de Conférence, Université de Strasbourg
Moreno Andreatta	Directeur de Recherche, Université de Strasbourg

RAPPORTEURS :

Isabelle Bloch	Professeure, Sorbonne Universités
Carlos Agon	Professeur, Sorbonne Université

AUTRES MEMBRES DU JURY :

Franck Hétroy-Wheeler	Professeur, Université de Strasbourg
Louis Bigo	Professeur, Université de Strasbourg

Thèse

INSTITUT DE RECHERCHE MATHÉMATIQUE AVANCÉE

UMR 7501

Strasbourg

présentée pour obtenir le grade de docteure de
l'Université de Strasbourg
Spécialité MATHÉMATIQUES

Victoria Callet

Modélisation topologique des structures et processus musicaux

Soutenue le 15 décembre 2023
devant la commission d'examen

Pierre Guillot, directeur de thèse
Moreno Andreatta, directeur de thèse
Isabelle Bloch, rapporteuse
Carlos Agon, rapporteur
Louis Bigo, examinateur
Franck Hétroy-Wheeler, examinateur

https://irma.math.unistra.fr

cnrs

IRMA
Institut de Recherche
Mathématique Avancée

Université
de Strasbourg

Modélisation topologique des structures et processus musicaux

Victoria Callet

Remerciements

Je tiens à m'excuser pour la longueur de ces remerciements, ce n'est ni pour gagner quelques pages en français ni parce que j'ai mal lu la consigne. On m'a dit de faire quelque chose qui me ressemble, et bien des personnes ont contribué à rendre ces quelques années incroyables, alors c'était vraiment important pour moi de les mentionner. Si vous le souhaitez, vous pouvez bien sûr passer directement à la partie qui vous concerne.

J'aimerai tout d'abord remercier chaleureusement mes directeurs de thèse, Pierre Guillot et Moreno Andreatta, sans qui tout cela n'aurait tout simplement pas été possible. Je remercie Pierre pour son soutien depuis le tout début, depuis le premier cours de deuxième année jusqu'à cette thèse en passant par Galois, un peu d'homologie et quelques polynômes symétriques. Merci de m'avoir présenté Moreno, grâce à qui j'ai pu découvrir ce monde merveilleux que sont les mathématiques et la musique. Merci à toi pour tous les exposés, les rencontres, les séminaires et Atlanta. Je vous remercie tous les deux pour votre soutien, votre écoute, vos relectures et vos conseils durant ces trois années, pour m'avoir proposé ce sujet de thèse et m'avoir donné cette chance incroyable.

Je remercie également les membres de mon jury pour leur temps et leur implication dans ma thèse. Merci à mes rapporteurs Isabelle Bloch et Carlos Agon pour leur relecture et leurs conseils, merci à mes examinateurs Louis Bigo et Franck Hétroy-Wheeler pour l'intérêt porté à mes travaux. Merci de vous être déplacés pour assister à ma soutenance, merci également pour les échanges que nous avons pu avoir durant mon doctorat.

Ce fut pour moi une chance incroyable de faire partie du monde de la recherche, et je tiens à remercier tous les membres de l'IRMA de m'avoir permis de le découvrir. Merci en particulier aux membres de l'équipe Algèbre, Représentation et Topologie, merci pour les séminaires et pour les GDR. J'aimerai également remercier les membres de mon comité de suivi, Pierre Baumann, pour ses conseils, son soutien et particulièrement ses cours de Master sans lesquels le confinement aurait eu une toute autre saveur, et Raphaël Côte, pour son écoute, ses conseils et son soutien durant ma thèse. Merci également aux membres de l'administration, plus particulièrement merci à Jessica et à Delphine pour leur gentillesse, leur patience et leur aide si précieuse, que ce soit dans l'organisation d'évènements ou simplement pour un déplacement. Je souhaite aussi remercier les membres de l'UFR math-info qui m'accueillent depuis plus de neuf ans, merci à Vincent Blanloeil, en particulier pour ses cours qui m'ont tant fait aimer l'algèbre, ainsi que Michael Gutnic, pour son écoute et son soutien depuis des années. Je tiens également à dire un grand merci à tous les enseignants géniaux que j'ai eus durant ma scolarité à Strasbourg et qui m'ont inspirée, merci à Charles Frances, Loïc Teyssier, Vincent Blanloeil, Pierre Guillot, Norbert Schappacher, Raphaël Côte, Nalini Anantharaman, Rao Bopeng, Nicolas Juillet, Giueseppe Ancona, Christine Vespa, Dragos Fratila, Pierre Baumann et Frédéric Chapoton. Un peu plus loin, merci à Christophe Del Degan pour m'avoir fait aimer les mathématiques il y a si longtemps. Merci à vous tous, je ne serais sans doute pas là où je suis aujourd'hui si je ne vous avais pas rencontrés.

Je souhaite également remercier toutes les personnes avec qui j'ai pu échanger sur mes travaux de recherche. Merci à toute la communauté « maths et musique », à tous les membres de l'IRCAM pour m'avoir accueillie quelques jours chez eux, merci particulièrement à Carlos et Gonzalo pour votre gentillesse et nos discussions. Je remercie également tous les invités de notre cher séminaire « mathémusique », et merci tout particulièrement à Alexandre d'avoir permis sa mise en place, merci pour ta disponibilité, tes conseils et tes coups de téléphone, sans quoi je n'y serais jamais arrivée. Merci à toute l'équipe d'organisation du MCM de m'avoir donné la chance de participer à cette incroyable aventure. Et merci à Ismail pour nos échanges, sur les mathématiques comme sur les mangas, merci pour ton soutien et ton accueil à Amiens.

J'aimerai également remercier toutes les personnes qui m'ont permis de m'impliquer dans les évènements filles et maths durant mon doctorat, en particulier merci à Laura et Marie de m'avoir fait entrer dans le RJMI, et merci à toutes celles et ceux que j'ai sollicités pour cet évènement, Antoine, Clarence, Brieuc, Adam, Thomas, Roméo, Céline, Romain, Rym, Romane, Suzanne, Morgane, Clémentine, Nathalie, Ana, Caroline, Christine, Ségolen, Laure, Anna, Solène... et qui ont toujours répondu présent. Merci aussi à la nouvelle vague, Céline, Roxana, Louise et Claire, le RJMI 2024 sera pour sûr une édition incroyable. Je tiens particulièrement à remercier Laurène, sans qui les éditions 2022 et 2023 auraient sans aucun doute bien moins fonctionné, merci pour ces moments si intenses, et bien sûr merci pour les pizzas et les madeleines. Enfin, je souhaite remercier toute l'équipe des Cigognes 2023 pour m'avoir permis de vivre cette aventure merveilleuse, pleine de rencontres folles et de moments hors du temps. Merci particulièrement à Clémentine d'être une si belle personne, merci à Pierre pour les rires et les sirops, merci à Samuel, Marie, Anne, Séréna, et toutes les (jeunes) Cigognes. Le doctorat n'aurait pas été le même sans tous ces évènements qui comptent tant à mes yeux, et je suis heureuse d'avoir pu y contribuer à mon échelle.

La thèse a été pour moi un épanouissement professionnel, mais avant tout une aventure humaine folle pleine de rencontres merveilleuses. Je tiens d'abord à remercier mon BFF Clarence, présent depuis le premier jour de ce M2 recherche, merci pour la boule à neige, pour les soirées gaufres-tiramisu, le crémant parfois un peu (trop) tôt et de ne jamais refuser une battle sur la Tribu de Dana, merci pour les cours particuliers à l'escalade et pour tous les fous rires. Merci Brieuc pour ton amitié, merci d'être toujours à l'écoute et de ne jamais refuser un café, merci pour toutes nos discussions, pour tes conseils et ton soutien sans faille, tes cours particuliers sur les ondes, merci pour l'escalade et les parties d'échecs. Merci Roméo d'être un super ami, merci pour toutes les discussions cinéma et sur bien d'autres sujets, pour les parties de coinches sans fin et les parties d'échecs et merci d'être toujours là pour faire discuter nos peluches (et désolée de ne pas avoir pu écrire vingt paragraphes). Merci Thomas d'avoir toujours le mot pour nous faire rire, pour tes adages farfelues, merci d'avoir été un examinateur exceptionnel pour le pass escalade, merci pour les séances de cinéma au deuxième rang (et parfois au rang G), pour les concerts privés du DJ Coconut et les impros sur Colonel Reyel. Merci Clément, mon irremplaçable Fatallemand sans qui Vitoo Ria ne serait rien, merci pour les chansons, les Bunny's, les fous rires et les girafes. Merci Basile pour les parties d'échecs et de tarots, les pâtes au pesto que j'espère goûter un jour, pour les GDR et leur Gala Dinner, merci d'être toujours là pour finir le vin sur les autres tables. Merci Raoul d'égayer notre bureau avec ta joie de vivre, ta bonne humeur et tes costumes parfois étranges, merci pour tes mots plus ou moins longs, pour les livres, les peluches et les discussions. Merci Céline d'être la meilleure cobureau B7, merci pour nos discussions qui ne sont jamais du commérage, merci pour l'escalade, les Apérols, les RJMI, les pulls doctorants, et merci d'être une duo irremplaçable sur The Climb. Merci Yohann et Thibault pour votre soutien depuis le début de cette thèse, pour vos conseils si précieux et toutes nos discussions café, et surtout merci d'avoir enrichi nos soirée de votre culture musicale. Merci Alexander d'avoir été un si gentil cobureau à notre arrivée, merci Vincent d'avoir accepté de faire parti du bureau des B7 et pour toutes nos chouettes discussions, merci Robin (même

si Sélestat c'est mieux et qu'on dit une Bretzel), merci Romain (depuis bien des années), merci Rym, Romane, Paul, Suzanne, Mickaël, Guillaume, Florian, Alex, Thomas A, Colin, Léopold, Jean-Pierre, Tom, Adam, Nicolas, Victor, Roxana, Claire, Louise, Jules,... Merci d'être passés nous voir en 117, que ce soit pour coincher, pour nous voler du café, pour passer l'été ou juste pour discuter.

Il faut être fou pour penser que les doctorants en mathématiques ne font que travailler, ils sont bien plus que ça et j'aimerais rendre hommage à tout ce qui a contribué à rendre ces quelques années vraiment belles et inoubliables. Un grand merci à toute la Team Cinéma, Antoine bien sûr, Thomas, Clarence, Roméo, Basile, Raoul, Rym, des (pires) Oscars aux (meilleurs) films français en passant par des avant-premières toujours plus gênantes, j'ai toujours adoré aller me divertir avec vous. Merci aussi pour les séances spéciales inoubliables du mardi soir, et d'avoir accepté de projeter Nos Jours Heureux au milieu de Quentin Dupieux et Massacre à la tronçonneuse. Merci à tous les grimpeurs hors pair que vous êtes, Clarence, Thomas S., Brieuc, Céline, Lépolod, Thomas A., Tom, Alex, Adam, Raoul, Roxana, merci de m'avoir fait découvrir ce sport, merci pour tous vos conseils si précieux, et merci de faire vivre Tom Cruise qui nous inspire à chaque séance. Merci aux coincheurs de tous les instants, Antoine, Roméo, Brieuc, Romane, Robin, Victor, Paul, Jean-Pierre, Basile, Florian (pour les tarots aussi), et merci à Basile et Adam de m'avoir fait revivre une année de compétitions d'échecs avec ce bon vieux Daniel, je trinque à notre trophée de troisième place en D1 (et à Basile l'invaincu) ! Merci pour les Comptoirs, les Télégraphes, les RA, les pauses cafés et les Chariots. Merci à vous tous les doctorants de Strasbourg, vous avez fait de ma thèse une aventure bien plus belle et plus riche que tout ce que j'aurais pu imaginer. J'ai appris tellement à vos côtés, et mon cœur est plein de souvenirs que je n'oublierai jamais.

J'aimerai remonter un peu plus loin, car la route des mathématiques a été longue et que j'y ai croisé des gens formidables. Quentin, les mots ne suffiraient certainement pas à te dire combien je te suis reconnaissante d'être présent depuis si longtemps, je ne serais probablement pas arrivée là où je suis si tu ne m'avais pas soutenue, si tu n'y avais pas cru avec moi. C'est un sacré chemin que nous avons parcouru, depuis les révisions rue Baldner (probas ou géométrie ?) en passant par quelques raclettes et une Masterclass. Merci d'être la personne que tu es, merci pour tout ce que tu as fait pour moi et ce que tu continues de faire, merci pour la musique, le MAThEOR, le Graspop, les Fingerspiels, les colloques de lettres, les champs de blé et les autoroutes, merci pour les fous rires et les pleurs, merci pour tout. Je crois qu'on peut être fiers de nous. So say we all. Je souhaite également remercier toutes les personnes qui m'ont accompagnée dans les mathématiques (et bien au-delà) durant cette folle aventure universitaire : merci à mon cher Renaud de m'avoir laissée m'assoir à côté d'un Lorrain à ce TD de calcul différentiel, merci pour notre périple à Lille et ces fous rires sur des parkings Belges, merci pour tous ces moments que nous avons passés ensemble et qui sont des souvenirs si précieux pour moi. Merci Morgane d'être une si belle personne, merci pour ta recette de cookies sur Sheryfa Luna, merci pour toutes nos discussions qui me sont si chères et d'être toujours à l'écoute. Merci Guillaume de toujours nous faire rire (et merci pour Jennie), merci Romain d'avoir rendu la compote de pomme unique, merci Chloé pour ton amitié depuis des années. Et enfin, merci à Céline d'être là depuis le tout début, merci de m'avoir demandé si j'avais aussi passé un Bac S lors de ce tout premier cours de Licence, merci d'avoir ponctué mon doctorat de quelques boucles d'oreilles, et merci tout simplement d'être la personne que tu es, merci d'être une super amie.

Je voudrai également glisser un petit mot à ma Dream Team, merci à Margaux, mon agent Pam de rêve, ma huitième, ma colloc', je ne te remercierai jamais assez d'avoir écrit "Spé Maths" sur la fiche de vœux en terminale. Merci à Coralie, Virginie et Julia d'être les copines les plus chouettes du monde, merci pour tous nos diners et piques-niques, toutes nos randonnées parfois un peu ratées, pour nos soirées jeux, toutes nos belles discussions et nos folles années passées sur ce banc. J'ai énormément de chance de vous avoir rencontrées, merci pour votre amitié depuis plus de dix ans, merci d'être toujours là.

J'ai eu la chance de grandir dans une maison qui a vu passé tellement de films, j'y ai appris des centaines de répliques, et je n'oublierai jamais que ce qu'il y a d'important dans la vie c'est pas les affaires, c'est pas l'argent, c'est la famille. Je ne serais jamais arrivée là où je suis sans le soutien incommensurable de la mienne. Papa, Maman, je vous serai toujours infiniment reconnaissante de m'avoir aidée à me construire, de m'avoir toujours soutenue dans mes choix et de m'avoir permis d'être la personne que je suis devenue. C'est peu de mots pour décrire tout ce que vous avez fait et continuez de faire pour moi, mais sans vous je n'y serais jamais parvenue. Mamie, j'aurai tellement aimé que tu sois là, je ne sais toujours pas qui a inventé les mathématiques, mais au fond, c'était pas si mal, non ? Merci papy pour tous les restaurants qui ont ponctué ma thèse et bien avant, et merci Marie-Claire d'avoir ouvert la voie à cette discipline dans notre famille. Merci aussi à ma belle-famille d'être venue assister à ma soutenance, et pour tous les moments passés ensemble. Je glisse un petit mot pour mon chien Dobby, qui a été le premier spectateur de chacun de mes exposés, et qui me soutient chaque jour par sa bienveillance et sa joie de vivre. Et bien sûr, je souhaiterais dire un mot pour chacun mes frères et sœurs, bien que cela ne suffira jamais à décrire tout ce qu'ils et elles représentent pour moi. Merci Jonas pour tes conseils musicaux depuis si longtemps, pour ton soutien et l'intérêt que tu portes à mes travaux, merci Antoine pour tes conseils sportifs et échiquéens et d'avoir toujours su garder ton âme d'enfant, merci Anna d'avoir accepté de grimper sur des prises et des marches chaque semaine, d'avoir couru sur l'eau sous les étoiles et d'être la meilleure Travolta, merci Elena pour toutes les surveillances d'amphi, nos promenades shopping qui se finissent parfois en concert un peu fou et d'être la meilleure Vaiana en chute libre, merci Samuel d'être toujours là pour un petit Knight and Day, d'être à la fois le mec, Renaud Le Pic et surtout une Schnouf en or, merci Paul pour tous nos skypes matinaux qui ont embelli mes journées de travail et pour tous tes conseils irremplaçables sur Pokémon, et merci Louis pour tes chansons parfois un peu folles mais toujours justes, tes blagues et surtout pour ton rire et ta joie de vivre à toute épreuve. Vous êtes mes fleurs, mon rayon de soleil. Merci Papa et Maman d'avoir créé la plus belle des tribus, et cette maison pleine de lumière dans laquelle on peut toujours se réfugier pour rire, pleurer, chanter ou simplement discuter. Merci d'être des parents merveilleux. Du fond du cœur, merci.

Mes derniers mots s'adressent à la plus belle et la plus douce personne que je connaisse, qui est présente chaque jour depuis plus de six ans. Merci de toujours croire en moi, de me soutenir, de m'encourager et de toujours m'aider à me relever. Merci d'avoir accepté de construire ensemble ce cocon dans lequel on se sent si bien. Merci d'être une personne exceptionnelle, sans qui je ne serais jamais arrivée au bout de cette aventure. Antoine, la route avec toi est tous les jours plus belle. Merci pour tout. Infiniment. A×♥V.

My soul is painted like the wings of butterflies,
Fairy tales of yesterday will grow but never die,
I can fly, my friends !

Brian May, Freddie Mercury
Queen, *The Show Must Go On* (1991)

Résumé

Cette thèse propose des contributions dans différents problèmes issus du domaine de l'analyse topologique de données musicales : les objets étudiés sont des partitions représentées symboliquement par des fichiers MIDI, et les outils utilisés sont la Transformée de Fourier Discrète et l'homologie persistante. Le manuscrit se découpe en trois parties : les deux premières sont consacrées à l'étude des objets mathématiques précédemment cités et à la mise en place du modèle. Plus précisément, la notion de DFT introduite par Lewin est généralisée au cas de la dimension deux, en explicitant le passage d'une mesure d'un morceau à un sous-ensemble de $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, ce qui conduit naturellement à une notion de métrique sur l'ensemble des mesures via leurs coefficients de Fourier. De cette construction naît un nuage de points auquel est associé le complexe filtré de Vietoris-Rips, et par suite une famille de codes-barres donnés par l'homologie persistante. Cette approche permet également de généraliser des résultats classiques tels que le lemme de Lewin ou encore le théorème de l'hexachorde de Babitt. La dernière partie de cette thèse est consacrée aux applications musicales du modèle basé sur la DFT : la première expérience consiste à extraire des codes-barres provenant de partitions artificiellement construites, telles que des gammes ou des accords plaqués. Cette étude mène en particulier à l'harmonisation de chansons que l'on réduit à leur mélodie et leur grille d'accords, ce qui permet entre autre de définir les notions de graphe et de complexité d'un morceau. L'homologie persistante se prête également au problème de la classification automatique du style musical, qui sera traité ici sous le prisme de descripteurs symboliques donnés par des statistiques calculées directement sur les codes-barres. Enfin, la dernière application propose un encodage des mesures basé sur la distance de Hausdorff conduisant à l'étude des textures musicales.

Abstract

This thesis proposes contributions to various problems in the field of topological analysis of musical data: the objects studied are scores represented symbolically by MIDI files, and the tools used are the discrete Fourier transform and persistent homology. The manuscript is divided into three parts: the first two are devoted to the study of the aforementioned mathematical objects and the implementation of the model. More precisely, the notion of DFT introduced by Lewin is generalized to the case of dimension two, by making explicit the passage of a musical bar from a piece to a subset of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, which leads naturally to a notion of metric on the set of musical bars by their Fourier coefficients. This construction gives rise to a point cloud, to which the filtered Vietoris-Rips complex is associated, and consequently a family of barcodes given by persistent homology. This approach also makes it possible to generalize classical results such as Lewin's lemma and Babitt's Hexachord theorem. The last part of this thesis is devoted to musical applications of the model: the first experiment consists in extracting barcodes from artificially constructed scores, such as scales or chords. This study leads naturally to song harmonization process, which reduces a song to its melody and chord grid, thus defining the notions of graph and complexity of a piece. Persistent homology also lends itself to the problem of automatic classification of musical style, which will be treated here under the prism of symbolic descriptors given by statistics calculated directly on barcodes. Finally, the last application proposes a encoding of musical bars based on the Hausdorff distance, which leads to the study of musical textures.

Contents

Introduction (Français)

Cette thèse s'intéresse à différents problèmes d'analyse musicale vus sous le prisme de la topologie algébrique. Plus précisément, notre but est de faire intervenir l'homologie persistante dans le cadre de l'analyse topologique de données appliquée à des représentations symboliques de partitions de musique, à savoir des fichiers MIDI. Les travaux que nous présentons ici s'inscrivent donc dans l'axe de recherche « traitement de l'information musicale à partir de représentations symboliques » de la communauté Music Information Research (voir [46]).

Dans cette introduction, nous proposons de dresser un état de l'art du domaine, puis nous donnons la problématique générale et présentons l'organisation du manuscrit.

Historique

✯ Partitions, fichiers MIDI et complexes simpliciaux

En analyse symbolique, le support de travail qui sert de représentation de morceaux de musique est donné par les fichiers MIDI. Un fichier MIDI (*Musical Instrument Digital Interface*) est une liste des évènements musicaux qui se produisent dans une partition donnée et qui sont présentés sous la forme d'une suite de messages symboliques. Ces données correspondent à des notes qui peuvent être organisées temporellement de la façon suivante :

$$(onset, length, pitch).$$

La hauteur (*pitch*) d'une note est encodée en *midicent*, à savoir des nombres entiers codés sur 7 bits permettant ainsi de travailler sur plus de dix octaves ($C_{-1} = 0$ et $G_9 = 127$ sur l'échelle chromatique tempérée). Le temps (*onset* et *length*) est quant à lui encodé à l'aide de *ticks* selon la correspondance usuelle suivante :

$$\text{𝅝} = 1920,\ \text{𝅗𝅥} = 960,\ \text{♩} = 480,\ \text{♪} = 240,\ \text{𝅘𝅥𝅯} = 120,\ \text{𝅘𝅥𝅰} = 60$$

Une vélocité est également associée à chaque note pour préciser l'intensité avec laquelle celle-ci est jouée. De telles suites de messages symboliques sont souvent représentées via le « Piano Roll », comme l'illustre la figure 1 (voir [49]).

Un fichier MIDI est donc une suite de messages qu'il est également possible de traiter par le biais de langages de programmation simples tels que Python. Dans cette thèse, nous utiliserons le langage Sage ([51]), qui s'apparente à ce dernier, et nous proposons quelques extraits de codes dans l'annexe A. de ce manuscrit. A noter qu'un fichier MIDI ne contient pas de son numérique, et que l'étude de ce type de données se fait par opposition à l'étude audio et au traitement du signal. Par ailleurs, l'une des problématiques de cette thèse fut centrée sur la recherche et le regroupement de fichiers MIDI sous la forme d'une base de données que nous présentons dans l'annexe B. de ce document. Cette base de données est disponible sur la page web dédiée :

`https://math-musique.pages.math.unistra.fr/midi.html`.

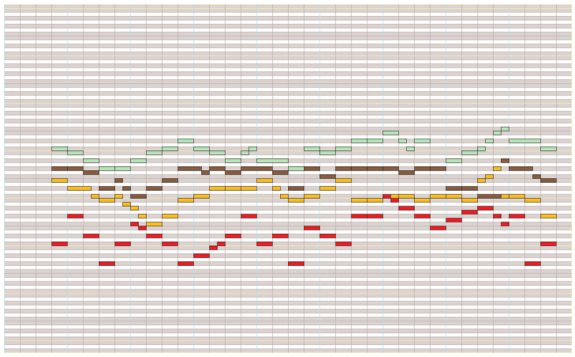


Figure 1: Piano Roll d'un fichier MIDI (figure extraite de [49]).

L'étude des fichiers MIDI fournit en outre une représentation fidèle d'un morceau de musique et a l'avantage de se modéliser aisément de façon géométrique, plus précisément au moyen de complexes simpliciaux. En effet, depuis une dizaine d'années, une telle modélisation symbolique est devenue un axe d'étude courant dans la communauté MIR, et l'on peut citer entre autre la thèse de L. Bigo présentée en 2013 ([11]) ainsi que les articles publiés à la suite de celle-ci ([15],[16]) qui constituent les premières contributions sur le sujet. Dans ces travaux, le principe consiste essentiellement à représenter des collections d'accords par des complexes simpliciaux dont les sommets sont donnés par des classes de hauteurs, comme l'illustre la figure 2. Parmi les applications proposées, on retrouve une représentation simpliciale du Tonnetz d'Euler, dont nous reprendrons la définition dans ce manuscrit (voir chapitre 5, section 5.2). Cette modélisation permet en particulier de visualiser un morceau de musique via sa représentation simpliciale en le projetant en deux dimensions dans le Tonnetz. En particulier, cette étude donnera naissance aux notions de trajectoire et de *compliance* d'une pièce, notions toutes deux introduites dans la thèse de L. Bigo. En parallèle, A. Spicher et ce dernier ont développé le locigiel « HexaChord » permettant de visualiser la trajectoire d'un fichier MIDI dans un Tonnetz choisi (voir [16]).

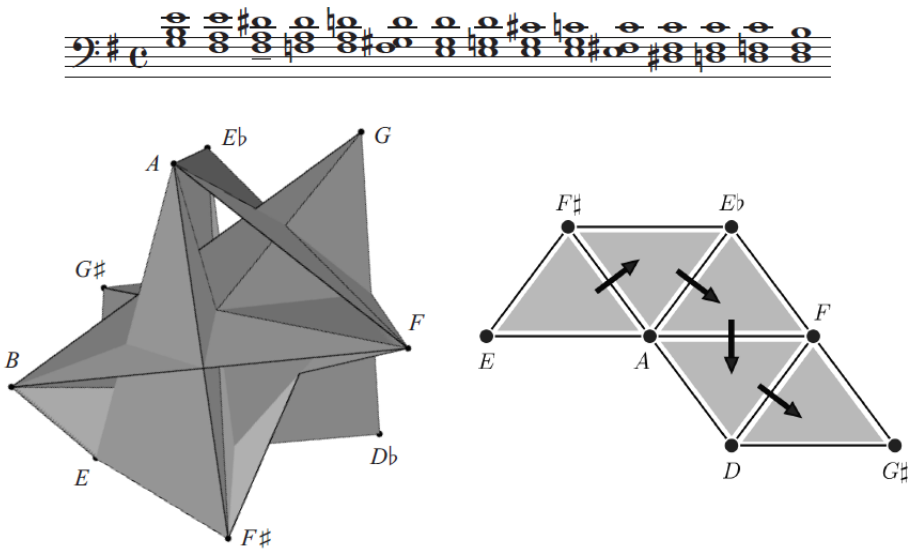


Figure 2: Les quinze premiers accords du Prélude Op. 28 No. 4 de F. Chopin. A gauche, sa représentation simpliciale. A droite, un chemin qui représente l'ordre des accords dans une région du complexe (figure extraite de [15]).

Cependant, cette représentation géométrique d'un morceau de musique reste incomplète compte tenu de la perte d'informations (notamment temporelle) qu'elle induit. En effet, un complexe simplicial représente une « photographie » d'un fichier MIDI à un instant donné, mais la notion de temps qui semble pourtant centrale dans l'étude d'un morceau de musique est laissée de côté afin de s'intéresser essentiellement à l'organisation des hauteurs. C'est ainsi que les notions de persistance et de filtration entre en jeu.

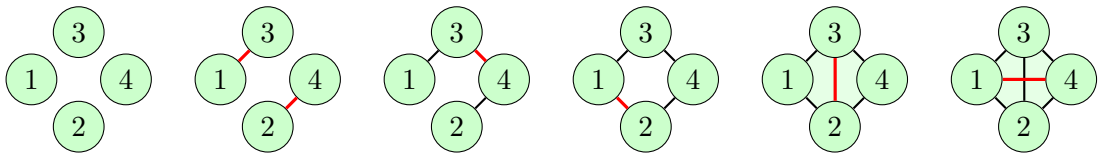
★ Filtration, persistance et premiers codes barres

L'idée d'étendre le procédé présenté dans le paragraphe précédent à l'homologie persistante est apparue naturellement dans la modélisation de morceaux de musique : en effet, si un complexe simplicial fournit une représentation géométrique d'une pièce à un instant donné (par exemple en associant un sommet à une classe de hauteur), la perte de l'information temporelle semble non négligeable. En ce sens, une filtration de complexes simpliciaux permet de contourner ce problème en ajoutant une dimension supplémentaire. Comme premiers travaux de recherche face à cette nouvelle problématique, on peut citer l'article de W. A. Sethares écrit en 2010 (voir [47]) et notamment la thèse de M. G. Bergomi présentée en 2016 (voir [8]).

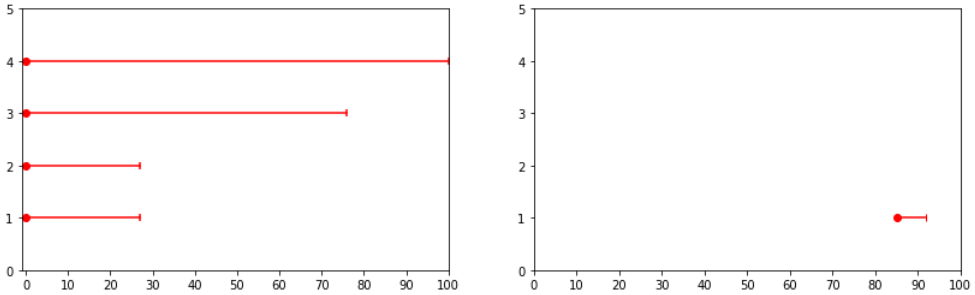
Le principe fondamental de l'homologie persistante est de généraliser le calcul de l'homologie simpliciale d'un complexe à celui d'une filtration, soit une suite de complexes imbriqués les uns dans les autres, et d'observer l'évolution des classes d'homologie à travers des diagrammes appelés codes-barres. Un code-barres est un graphique sur lequel sont représentées les durées de vie des classes d'homologie à degré fixé dans une filtration donnée. Ces durées sont matérialisés par des « barres » dont le début et la fin correspondent respectivement à la « naissance » et à la « mort » d'une classe. Ainsi, si $\mathcal{F} = (K^i)_i$ est une filtration (finie) de complexes simpliciaux et $\mathrm{BC}_d(\mathcal{F})$ le code-barres associé à un degré fixé d, on peut simplement décrire ce code-barres par un ensemble de points dans $\mathbb{N}^2$:

$$\mathrm{BC}_d(\mathcal{F}) = \{(b_i, d_i) \mid b_i \text{ début de la classe } i,\ d_i \text{ fin de la classe } i\} \subset \mathbb{N}^2.$$

La figure 3 donne un exemple de filtration de complexes simpliciaux et de la famille de codes-barres associée (en degrés 0 et 1).



(a) Une filtration de complexes simpliciaux.



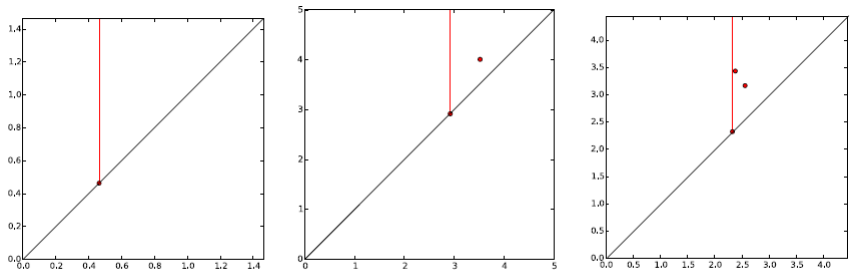
(b) La famille de codes-barres associée à la filtration 3a en degré 0 (gauche) et degré 1 (droite).

Figure 3: Filtration de complexes simpliciaux et codes-barres.

En analyse topologique de données, un des objectifs principaux est d'étudier un objet fixé (dans notre cas un morceau de musique) à travers ses caractéristiques topologiques. Pour ce faire, on en extrait un nuage de points dont on construit un complexe filtré, puis on calcule l'homologie persistante associée, que l'on représente par une famille de codes-barres. En effet, ce sont ces représentations graphiques de l'homologie persistante qui nous aident à identifier les éléments topologiques caractéristiques de notre objet de départ : plus précisément, une classe qui « survit » durant une majeure partie de la filtration est un élément qui persiste et qui représente donc potentiellement une propriété topologique importante, tandis qu'une classe qui « meurt » rapidement représente du bruit et peut être ignorée. Ces notions de durée de vie se traduisent aisément par la longueur des barres dans le code-barres considéré.

Ainsi, l'homologie persistante permet d'associer une signature topologique à une partition de musique via les codes-barres. Dans la littérature, on retrouve souvent une représentation de ces graphiques sous la forme de diagrammes de persistance. Il s'agit de graphiques où les abscisses et ordonnées représentent respectivement la naissance et la mort des classes d'homologie, et où une classe qui persistent est une barre verticale. Par exemple, la figure 4 extraite de la thèse de Bergomi illustre des diagrammes de persistance obtenus sur une famille de morceaux de musique classique. Ces diagrammes ont été construits en plongeant des séquences musicales dans le Tonnetz suivant le travail effectué dans [15], et en utilisant ensuite la durée des notes dans le but de « déformer » le Tonnetz et ainsi de filtrer les séquences étudiées.

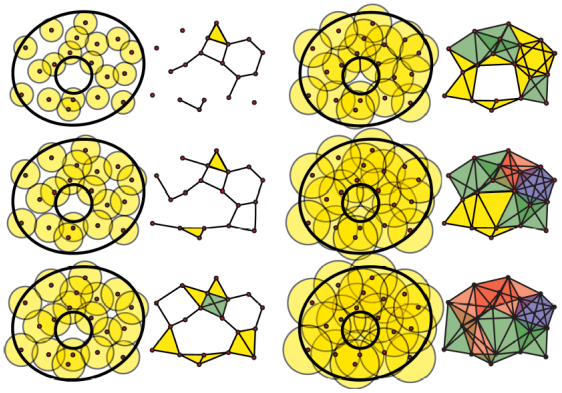
Parmi les travaux actuels d'applications de l'homologie persistante à l'analyse musicale essentiellement basés sur la thèse de Bergomi, on peut citer entre autre les articles [10], [14], [39] et [52].



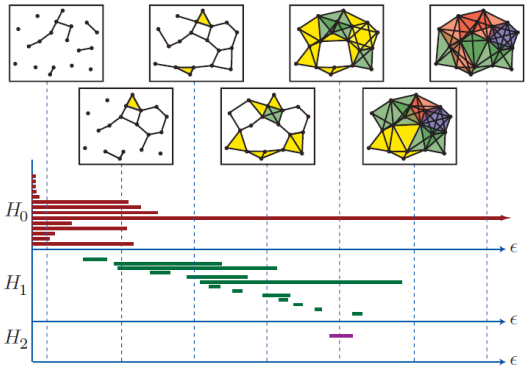
(a) Arabesque, Debussy. (b) Jeux d'Eau, Ravel. (c) Klavierstück I, Schönberg.

Figure 4: Les diagrammes de persistance en degrés 0 représentants les empreintes topologiques associées à trois compositions différentes (figure extraite de [8]).

Un autre travail réalisé quelques années plus tôt autour de l'analyse topologique de données musicales est celui de W. A. Sethares (voir [47]). Dans cet article paru en 2010, l'auteur ne propose pas de transformer une pièce en un complexe simplicial via les classes de hauteurs et ne mentionne pas le Tonnetz. En revanche, il propose de reconnaître trois structures musicales classiques - le cercle chromatique, le cercle des quintes et une représentation circulaire du rythme - dans différentes morceaux de musique. En mettant en place une distance sur les classes de hauteurs et les rythmes, l'auteur convertit un fichier MIDI en une matrice, lui permettant ainsi de passer à un nuage de points puis à un complexe simplicial filtré en appliquant la méthode de Vietoris-Rips (ou clique-complexes). Ce procédé consiste à ajouter des n-simplexes au nuage de points à mesure qu'un certain paramètre augmente. Plus ce paramètre est petit et plus les points sont séparés tandis qu'à l'inverse, plus ce paramètre est grand et plus le complexe obtenu est trivial et topologiquement équivalent à un seul point. L'intérêt est donc d'arrêter la filtration et d'analyser le complexe correspondant au « bon » paramètre, c'est à dire au moment où les caractéristiques topologiques sont correctement représentées. Cette méthode de construction est présentée dans l'article [27] de R. Ghrist ainsi que dans le chapitre 4 de cette thèse, et la figure 5 en illustre le procédé.



(a) Une filtration de Rips associée à un nuage de points représentant un anneau. Plus ϵ augmente, plus des cycles apparaissent et disparaissent. Lesquels d'entre eux sont réels et lesquels sont du bruit ?



(b) [en bas] Un exemple de codes-barres pour $H_*(R)$ dans l'exemple de la figure 5a. [en haut] Le rang de $H_k(R_{\epsilon_i})$ est égal au nombre d'intervalles du code-barres de $H_k(R)$ selon la ligne (en pointillés) $\epsilon = \epsilon_i$.

Figure 5: Filtration de Vietoris-Rips et codes-barres associés (figures extraites de [27]).

De cette façon, Sethares calcule des codes-barres sur des morceaux artificiellement construits dans un premier temps comme des gammes, puis sur une base de données de fichiers MIDI : il y retrouve les structures musicales attendues en étudiant essentiellement les degrés 0 et 1. A titre d'exemple, on peut citer la figure 6 extraite de [47] qui représente différents codes-barres obtenu à partir d'un Choral de Bach.

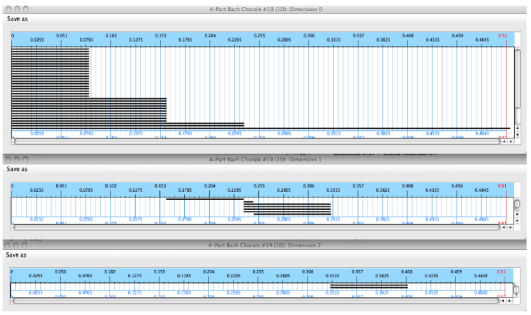


Figure 6: Codes-barres obtenus à partir du Choral de Bach No. 19 (figure extraite de [47]).

Les travaux de la dernière décennie effectués autour de l'homologie persistante appliquée à l'analyse musicale présentent les pistes de recherche actuelles, et toutes démarrent avec la même question fondamentale, qui constitue également la problématique de départ de cette thèse : comment représenter un morceau de musique par une filtration de complexes simpliciaux ? L'idée d'associer une classe de hauteur à un sommet semble intuitive mais a l'inconvénient de produire des complexes de très grande dimension (au moins douze si on ne se restreint qu'à une seule octave) et de laisser de côté les rythmes ainsi que les durées des notes. Projeter dans un Tonnetz ou encore filtrer en utilisant les durées sont des choix qui permettent d'affiner la première idée et qui semblent produire des résultats prometteurs en ce qui concerne la classification stylistique (voir le paragraphe suivant). Dans cette thèse, nous ferons le choix arbitraire de décrire une pièce de musique à travers ses mesures : pour une partition donnée, il existe un découpage naturel donné par la signature rythmique du morceau. Ce découpage permet ainsi de « zoomer » sur des morceaux de la partition, de la même façon qu'on découperait une image en plusieurs pièces d'un même puzzle. Chacune de ces pièces peut alors être décrite comme un sous-ensemble $\mathcal{B}$ de la forme :

$$\{n = (onset, pitch) \mid n \in \mathcal{B}\} \qquad (\star)$$

où les coordonnées *onset* et *pitch* vivent respectivement dans $\mathbb{Z}/t\mathbb{Z}$ et $\mathbb{Z}/p\mathbb{Z}$, avec t une unité de temps et p une unité de hauteur. Ainsi, une partition est définie comme un ensemble fini non-ordonné de mesures distinctes, et cette description permet d'associer un nuage de points à une partition, où chaque point est alors donné par une mesure contenant les informations rythmiques et fréquentielles des notes jouées.

★ Le problème de la classification automatique du style

Parmi les applications majeures de l'analyse topologique de données musicales, on retrouve bien souvent la question de la classification automatique du style, qui fera l'objet du chapitre 7 de cette thèse. L'article [10] de Bergomi en est un précurseur, et on peut citer à titre d'exemple la figure 7 sur laquelle on observe un dendrogramme obtenu à l'aide des diagrammes de persistance munis de la distance Bottleneck.

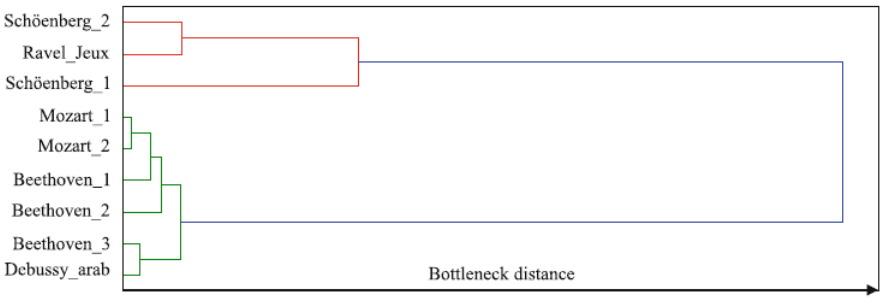


Figure 7: Dendrogramme basé sur la persistance de neuf pièces classiques et contemporaines (figure extraite de [10]).

Une question essentielle ici est donc celle de la distance sur l'ensemble des codes-barres : en effet, si ces derniers permettent d'associer une signature topologique à un morceau de musique, il semble nécessaire de définir une distance sur l'ensemble de ces objets si l'objectif est de comparer plusieurs pièces entre elles. Comme nous l'avons énoncé brièvement plus haut, la métrique la plus populaire sur les diagrammes de persistance est donnée par la distance Bottleneck (voir par exemple [10], [19], ou encore [32]). Si BC_1 et BC_2 sont deux diagrammes de persistantes (ou codes-barres), on pose $\phi : \mathrm{BC}_1 \to \mathrm{BC}_2$ une des applications qui, à chaque élément de BC_1, associe soit un élément de BC_2 si les codes-barres sont de même cardinaux, soit dans le cas contraire son projeté sur la diagonale $x = y$. La distance Bottleneck entre BC_1 et BC_2 est alors donnée par $(\star\star)$:

$$d_q(\mathrm{BC}_1, \mathrm{BC}_2) = \inf_{\phi \,:\, \mathrm{BC}_1 \to \mathrm{BC}_2} \left(\sup_{(b,d) \in \mathrm{BC}_1} \| \, (b,d) - \phi(b,d) \, \|_q \right) \qquad (\star\star)$$

La figure 8 extraite de [32] présente un exemple de calcul de la distance Bottleneck. Sur cette illustration, on constate bien que cette définition provient d'une interprétation géométrique des diagrammes de persistance, mais qui laisse de côté l'interprétation des codes-barres du point de vue de la durée de vie des classes d'homologie, et donc de la longueur des barres. Dans cette thèse, notre représentation de l'homologie persistante sera exclusivement donnée par des codes-barres, qui nous semblent en effet plus proches de l'interprétation en terme de persistance des classes, et donc bien de la longueur des barres. De ce fait, la distance Bottleneck n'apparaît pas comme

un outil de comparaison intuitif de tels graphiques, et nous proposons donc de comparer deux codes-barres en calculant des statistiques (moyenne, écart-type et entropie) sur ces ensembles, comme présenté dans l'article [39]. Ce point de vue nous permettra entre autre de produire des résultats de classifications prometteurs qui seront présentés dans le chapitre 7 de ce manuscrit.

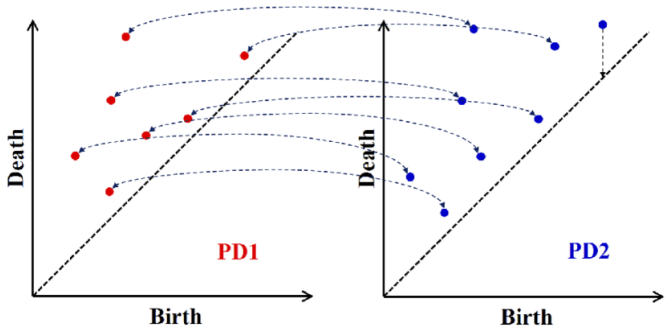


Figure 8: Description schématique de la manière de calculer la distance Bottleneck entre deux diagrammes de persistance (PD) (figure extraite de [32]).

Notons par ailleurs que le problème de la classification automatique du style et de distance entre les codes-barres n'est pas l'unique sujet auquel se prête l'homologie persistante dans le contexte de l'analyse musicale, et qu'il ne constitue qu'une seule des quatre applications proposées dans cette thèse. Par exemple, l'objectif de Sethares était essentiellement de retrouver, au moyen des codes-barres, des structures musicales dans des morceaux parfois artificiellement construits, comme simplement des gammes. Cette étude se rapproche des problèmes traités dans les chapitres 5 et 6 de ce manuscrit, dans lesquels notre objectif est d'analyser des codes-barres provenant de partitions construites à partir de gammes, de Tonnetze ou encore de morceaux harmonisés extraits de la musique Pop.

★ De l'analyse de données à la DFT

Associer une filtration de complexes à un morceau de musique nécessite d'en extraire un nuage de points, à savoir une collection de points dans un espace métrique. Comme mentionné dans les paragraphes précédents, nous appliquerons pour cela la méthode de Vietoris-Rips sur un nuage de points donné par les mesures d'une partition. Il reste donc à établir une distance entre nos sommets, et c'est à cet endroit qu'intervient la Transformée de Fourier Discrète (DFT).

La DFT est un outil mathématique qui permet de modéliser des séquences musicales simples, comme des gammes ou des collections de rythmes. Historiquement, elle a été introduite dans le contexte de l'analyse musicale en 1959 par D. Lewin (voir [35]), dont l'objectif était de formaliser la notion de contenu intervallique entre deux ensembles de classes de hauteur. Cette formalisation a conduit au lemme de Lewin, que nous généralisons dans cette thèse au cas de la dimension deux (voir théorème 2.2.4). Au début des années 2000, une série de travaux ont suivi le dernier article de Lewin sur le sujet ([36]), parmi lesquels on peut citer les problèmes liés à la notion d'homométrie (voir [37], [38]) ou encore de canons rythmiques (voir [3], [55]). Suite à cela, la DFT et ses applications à l'analyse musicale se sont réellement popularisées notamment grâce aux travaux d'E. Amiot et J. Yust (voir par exemple [4], [5], [54] et [56]).

Dans cette thèse, nous nous intéressons essentiellement à la modélisation d'une séquence musicale au moyen de la DFT : si $\mathcal{P} = (p_1, \ldots, p_N)$ est une suite de classes de hauteurs (ou de données rythmiques) dans $\mathbb{Z}/n\mathbb{Z}$, alors on peut lui associer une liste de n coefficients de Fourier $\big(\mathcal{F}_{\mathcal{P}}(0), \ldots, \mathcal{F}_{\mathcal{P}}(n-1)\big)$, où $\mathcal{F}_{\mathcal{P}}(x)$ est donné par la somme d'exponentielles complexes suivante

$$\mathcal{F}_{\mathcal{P}}(x) = \sum_{k\in\mathcal{P}} \exp\left(\frac{-2i\pi kx}{n}\right).$$

Ainsi, comparer deux séquences musicales données revient à comparer leur liste respectives de coefficients de Fourier. La figure 9 extraite du livre d'Amiot [4] illustre la comparaison entre une mélodie et son accompagnement dans un morceau de Bartok. Les séquences musicales sont modélisées via le module des coefficients de Fourier.

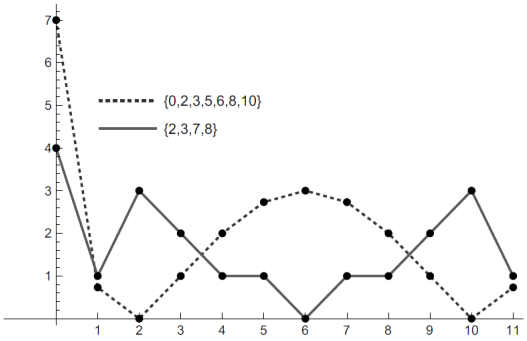


Figure 9: La DFT (en module) d'une mélodie et son accompagnement chez Bartok (figure extraite de [4]).

Comme nous l'avons mentionné précédemment, l'objectif principal de cette thèse est de modéliser des partitions en utilisant une description donnée par leurs mesures. Dans ce contexte, la DFT se révèle prometteuse car, en généralisant le modèle précédent au cas de la dimension deux, on obtient une représentation similaire à la notion de mesure évoquée en $(\star)$, et comparer deux telles mesures revient alors à comparer les matrices de coefficients de Fourier associées. Notre modèle est donc le suivant : une partition de musique est un ensemble fini non-ordonnés de mesures distinctes $\{\mathcal{B}_1, \dots \mathcal{B}_N\}$, et chaque mesure est un sous-ensemble de $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, où t est une unité de temps (par exemple $t = 8$ si on choisit la croche comme unité) et p une unité de hauteurs (par exemple $p = 12$ si on travaille dans une octave). On note $\mathfrak{B}_{t,p}$ l'ensemble de toutes les mesures de $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. Alors, on associe à $\mathfrak{B}_{t,p}$ la métrique donnée par la DFT en deux dimensions : si $\mathcal{B}$ et $\mathcal{B}'$ sont deux mesures dans $\mathfrak{B}_{t,p}$, alors leur distance est donnée par

$$\mathrm{d}_{\mathrm{DFT}}(\mathcal{B}, \mathcal{B}') = \sum_{x=1}^{t} \sum_{y=1}^{p} |\mathcal{F}_{\mathcal{B}}(x,y) - \mathcal{F}_{\mathcal{B}'}(x,y)|$$

où $\mathcal{F}_{\mathcal{B}}(x,y)$ est le (x,y)-coefficient de Fourier de $\mathcal{B}$ défini par

$$\mathcal{F}_{\mathcal{B}}(x,y) = \sum_{(k,l)\in\mathcal{B}} \exp\left(\frac{-2i\pi kx}{t}\right) \exp\left(\frac{-2i\pi ly}{p}\right).$$

Ainsi, $(\mathfrak{B}_{t,p}, \mathrm{d}_{\mathrm{DFT}})$ est un espace métrique, et une partition de musique devient un nuage de points, où chaque point est donné par un sous-ensemble de $(\mathfrak{B}_{t,p}, \mathrm{d}_{\mathrm{DFT}})$. Alors, grâce à la méthode de Vietoris-Rips présentée plus haut, on associe une filtration de complexes à un morceau de musique via le nuage de points ainsi construit, ce qui permet alors de calculer l'homologie persistante et d'en extraire une famille de codes-barres. Il s'agit du point de vue qui sera étudié tout au long de ce manuscrit. D'autre part, cette modélisation offre la possibilité de généraliser des résultats théoriques classiques sur la DFT en dimension deux, comme par exemple le lemme de Lewin sur le contenu intervallique que nous avons déjà évoqué, ou encore le théorème de l'hexachorde de Babbitt ([6]).

Organisation du document

Ce manuscrit se découpe en trois grandes parties : les deux premières sont consacrées à l'étude des objets mathématiques utilisés, à savoir l'homologie persistante et la transformée de Fourier discrète, tandis que la troisième est réservée aux différentes applications musicales.

Dans la première partie (I), nous généralisons la notion de DFT introduite par Lewin au cas de la dimension deux. Plus précisément, nous explicitons dans le chapitre 1 le passage d'une mesure à un sous-ensemble de $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, puis nous définissons la notion de DFT-distance sur l'ensemble des mesures. Dans le chapitre 2, nous proposons de généraliser certains résultats théoriques classiques sur la DFT tels que le lemme de Lewin sur le contenu intervallique (théorème 2.2.4) ou encore le théorème de l'hexachorde de Babbitt (théorème 2.3.1). Enfin, nous clôturons cette première partie par l'étude des isométries sur l'ensemble $(\mathfrak{B}_{t,p}, \mathrm{d}_{\mathrm{DFT}})$ des mesures de $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ (section 2.4).

La deuxième partie de cette thèse (II) est découpée en deux chapitres : dans le premier (3), nous faisons quelques rappels de théorie simpliciale, puis nous nous concentrons sur la définition de filtration, d'homologie persistante et plus particulièrement de codes-barres. Dans le chapitre suivant (4), nous définissons le modèle qui servira de support tout au long de cette thèse, à savoir la modélisation d'une partition comme nuage de points via ses mesures et la DFT-distance.

Dans la troisième et dernière partie de ce manuscrit (III), nous proposons quatre applications musicales de notre modèle. La première est consacrée au problème inverse, à savoir appliquer notre procédé d'extraction de codes-barres à des partitions artificiellement construites, telles que des gammes ou des Tonnetze. On retrouve alors des caractéristiques topologiques connues comme par exemple le tore sur le Tonnetz d'Euler (chapitre 5). Dans le chapitre 6, nous proposons d'étudier des morceaux de musique Pop harmonisés en réduisant uniquement à deux pistes, la mélodie et l'accompagnement. Ce procédé permet notamment d'associer un « graphe-type » et une « complexité » à un morceau donné. La troisième application constitue une partie essentielle de cette thèse, puisqu'on s'intéresse ici à la problématique de la classification automatique du style musical via l'homologie persistante (chapitre 7), où notre modèle basé sur la DFT semble produire des résultats prometteurs. Enfin, le chapitre 8 est consacré à une façon différente d'encoder les mesures d'un morceau via une autre distance donnée par la métrique de Hausdorff, qui fournit notamment des résultats satisfaisants sur la structure globale de certains morceaux de musique.

Nous clôturons ce manuscrit par une conclusion générale et des annexes consacrées aux codes utilisés pour notre travaux (Annexe A.) et à une présentation de notre base de données MIDI (Annexe B.).

Introduction (English)

This thesis focuses on various music-analytical problems seen through the prism of algebraic topology. More specifically, our aim is to apply persistent homology to topological data analysis with symbolic representations of musical scores such as MIDI files. The work we present here is therefore part of the "processing musical information from symbolic representations" research axis of the Music Information Research community (see [46]).

In this introduction, we present the state of the art in the field, then outline the general problem and present the organization of the manuscript.

Historical survey

✯ Scores, MIDI files and simplicial complexes

In symbolic analysis, the support for representing musical pieces and scores is provided by MIDI files. A MIDI file (Musical Instrument Digital Interface) is a list of musical events occurring in a given score, presented in the form of a sequence of symbolic messages. These data correspond to notes, which can be organized temporally as follows:

$$\text{(onset, length, pitch)}.$$

The pitch of a note is encoded in *midicent*, i.e. whole numbers encoded on 7 bits, enabling us to work on more than ten octaves ($C_{-1} = 0$ and $G_9 = 127$ on the tempered chromatic scale). Time (*onset* and *length*) is encoded using *ticks* according to the following usual correspondence:

$$\text{𝅝} = 1920,\ \text{𝅗𝅥} = 960,\ \text{♩} = 480,\ \text{♪} = 240,\ \text{𝅘𝅥𝅯} = 120,\ \text{𝅘𝅥𝅰} = 60$$

A velocity is also associated with each note to specify the intensity with which it is played. Such symbolic message sequences can be interpreted via the "Piano Roll", as illustrated in Figure 1 (see [49]).

A MIDI file is therefore a sequence of messages that can also be processed using simple programming languages such as Python. In this thesis, we will be using the Sage language ([51]), which is similar to Python, and we will be providing some extracts of codes in Appendix A. of this manuscript. Note that a MIDI file contains no digital sound, and that the study of this type of data is complementary to the study of audio and signal processing. In addition, one of the issues addressed in this thesis was the retrieval and grouping of MIDI files in the form of a database, which we present in Appendix B. of this document. This database is also available on the dedicated web page:

`https://math-musique.pages.math.unistra.fr/midi.html`.

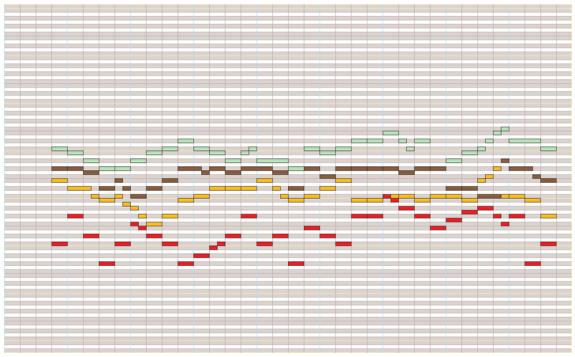


Figure 1: Piano Roll of an MIDI file (extract from [49]).

The study of MIDI files also provides a faithful representation of a piece of music, and has the advantage of being easily modeled geometrically, more precisely by means of simplicial complexes. In fact, over the last ten years, such symbolic modeling has become a common line of study in the MIR community: we can mention L. Bigo's thesis defended in 2013 ([11]) and the articles published following it ([15],[16]), which are the first contributions on the subject. In this paper, the principle essentially consists in representing collections of chords by simplicial complexes whose vertices are given by pitch-classes, as illustrated in Figure 2. Among the applications proposed, the authors provide a simplicial representation of Euler's Tonnetz, whose definition is summarized in this manuscript (see Chapter 5, Section 5.2). In particular, this modeling allows us to visualize a piece of music by its simplicial representation, by projecting it in two dimensions into the Tonnetz. In particular, this study gave rise to the notions of trajectory and compliance of a piece, notions both introduced in L. Bigo's thesis. At the same time, A. Spicher and L. Bigo developed the "HexaChord" software for visualizing the trajectory of a MIDI file in a selected Tonnetz (see [16]).

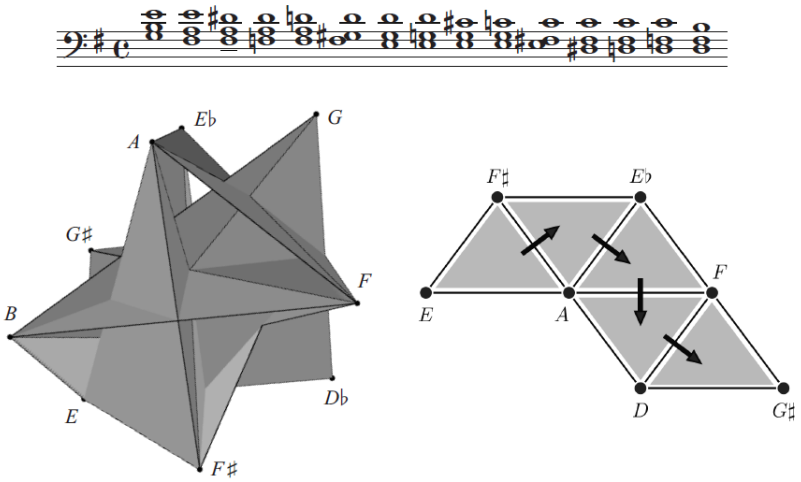


Figure 2: Fifteen first chords of Chopin's Prelude Op. 28, No. 4. On the left its simplicial representation. On the right, a path represents the order of chords in a region of the complex (extract from [15]).

However, this geometric representation of a piece of music remains incomplete, given the loss of (temporal) information that it induces. Indeed, a simplicial complex represents a "picture" of a MIDI file at a given moment, but the notion of time, which seems central in the study of a musical piece, is left aside in order to focus essentially on the organization of pitches. This is how the notions of persistence and filtration comes into play.

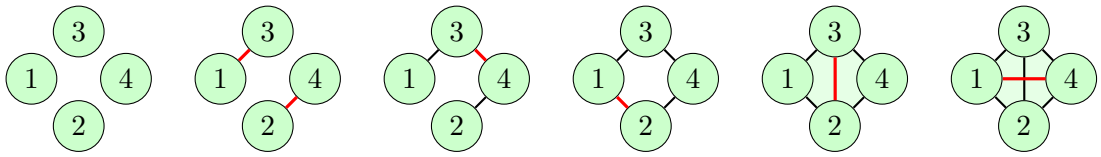
★ Filtration, persistence and first barcodes

The idea of extending the procedure presented in the previous paragraph to persistent homology came naturally in the modeling of musical pieces: indeed, if a simplicial complex provides a geometric representation of a piece at a given instant (for example, by associating a vertex with a pitch class), the loss of temporal information seems non-negligible. In this sense, a filtration of simplicial complexes allows us to get around this problem by adding an extra dimension. Initial research work on this new problem includes W. A. Sethares's paper written in 2010 (see [47]) and M. G. Bergomi's thesis defended in 2016 (see [8]).

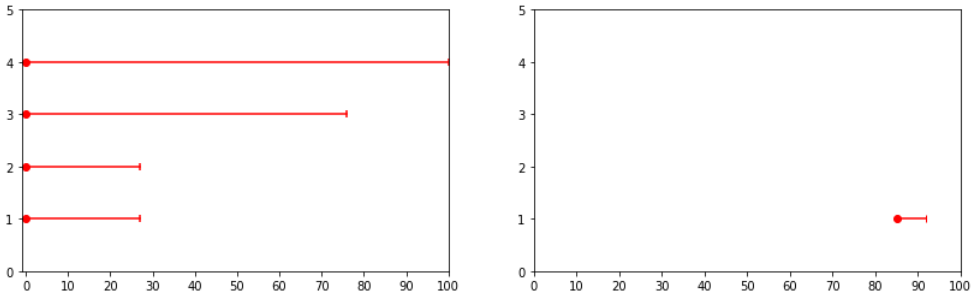
The fundamental principle of persistent homology is to generalize the calculation of the simplicial homology of a single complex to a whole filtration, i.e. a sequence of complexes nested within each other, and to observe the evolution of homology classes through diagrams called barcodes. A barcode is a graph on which are represented the lifetimes of homology classes at a fixed degree in a given filtration. These durations are represented by "bars", whose beginning and end correspond respectively to the "birth" and "death" of classes. Thus, if $\mathcal{F} = (K^i)_i$ is a (finite) filtration of simplicial complexes and $\mathrm{BC}_d(\mathcal{F})$ the associated barcode with a fixed degree d, we can simply describe this barcode by a set of points in $\mathbb{N}^2$:

$$\mathrm{BC}_d(\mathcal{F}) = \{(b_i, d_i) \mid b_i \text{ birth of class } i,\ d_i \text{ death of class } i\} \subset \mathbb{N}^2.$$

Figure 3 shows an example of simplicial complex filtration and its associated family of barcodes (in degrees 0 and 1).



(a) A filtration of a simplicial complex.



(b) The associated family of barcodes with 3a in degree 0 (left) and degree 1 (right).

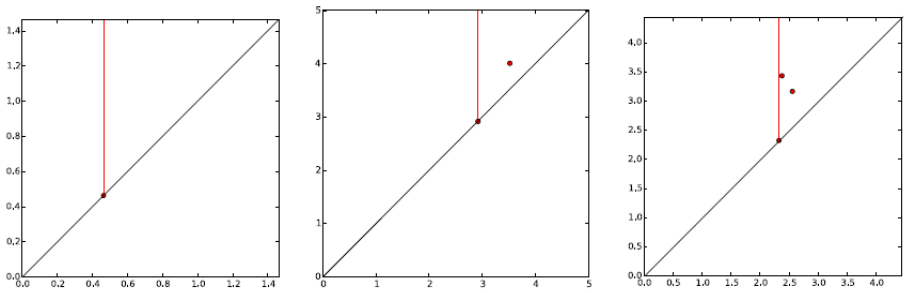
Figure 3: Filtration and barcodes.

In topological data analysis, one of the main objectives is to study a fixed object (in our case, a piece of music) through its topological features. To do this, we extract a point cloud from which we construct a filtered complex, then calculate the associated persistent homology, which we represent by a family of barcodes. Indeed, these graphical representations of the persistent homology help us to identify the characteristic topological elements of our starting object: more precisely, a class that "survives" for a major part of the filtration is an element that persists and therefore represents an important topological property, whereas a class that "dies" quickly represents noise and can be ignored. These notions of lifetime are easily translated into the length of the bars in barcodes.

In this way, persistent homology makes it possible to associate a topological signature with a musical score by means of barcodes. In the literature, these graphs are often represented as

persistence diagrams, which are graphs where the x-axis and y-axis represent respectively the birth and death of homology classes, and where a persisting class is a vertical bar. For example, Figure 4 from Bergomi's thesis illustrates persistence diagrams obtained on a family of classical music pieces. These diagrams were constructed by embedding musical sequences into the Euler's Tonnetz following the approach described in [15], and then using note duration to "deform" the resulting Tonnetz and filter the studied sequences.

Among current works on the application of persistent homology to music analysis essentially based on Bergomi's thesis, we can mention the articles [10], [14], [39] and [52].



(a) Arabesque, Debussy. (b) Jeux d'Eau, Ravel. (c) Klavierstück I, Schönberg.

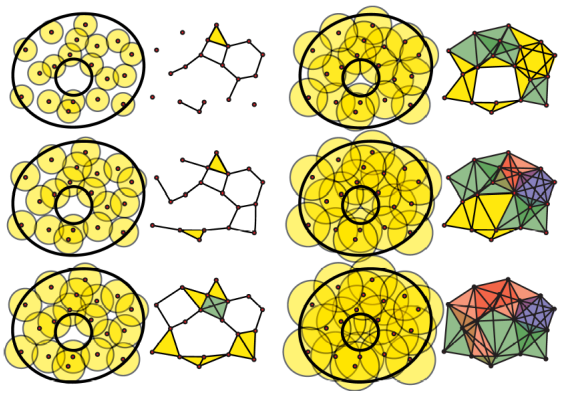
Figure 4: The 0-persistence diagrams representing the topological fingerprints associated to three different compositions (extract from [8]).

Another work carried out a few years earlier around the topological analysis of musical data is the article by W. A. Sethares in 2010 (see [47]). In this paper, the author does not propose to transform a piece into a simplicial complex via pitch-classes, nor does he mention the Tonnetz. Instead, he proposes to recognize three classical musical structures - the chromatic circle, the circle of fifths and a circular representation of rhythm - in different pieces of music. By setting up a distance on pitch-classes and rhythms, the author converts a MIDI file into a matrix, allowing him to pass to a point cloud and then to a filtered simplicial complex by applying the Vietoris-Rips (or clique-complex) method. This process consists of adding n-simplices to the point cloud as a certain parameter increases. The smaller the parameter is, the more separated the points are, while conversely, the larger the parameter is, the more trivial and topologically equivalent to a single point the resulting complex is. The point is therefore to analyze the complex for the "right" parameter, i.e. when the topological features are correctly represented. This construction method is presented in the article [27] by R. Ghrist and in Chapter 4 of this thesis, and Figure 5 illustrates the procedure. In this way, Sethares first calculates barcodes on artificially constructed pieces such as musical scales, then on a database of MIDI files: he finds the expected musical structures by essentially studying the 0 and 1 degrees. An example is Figure 6 from [47], which represents some barcodes obtained from a Bach Chorale.

During the last decade, work on persistent homology applied to music analysis presents the current lines of research, all raising the same fundamental question, which is also the starting point of this thesis: how can a piece of music be represented by a filtration of simplicial complexes? The idea of associating a pitch class with a vertex seems intuitive, but has the disadvantage of producing very high-dimensional complexes (at least twelve if we restrict the study to a single octave!) and leaving out the rhythms and durations of the notes. Projecting into a Tonnetz or filtering using durations are choices that help refine the first idea and seem to produce promising classification results (see next paragraph). In this thesis, we will make the arbitrary choice of describing a piece of music through its musical bars: for a given score, there is a natural division given by the meter of the piece. This division allows us to zoom in on parts of the score, in the same way as we would divide a picture into several pieces of the same puzzle. Each of these pieces can then be described as a $\mathcal{B}$ subset of the form

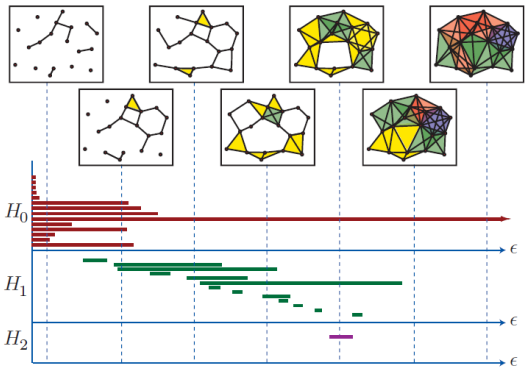
$$\{n = (onset, pitch) \mid n \in \mathcal{B}\} \qquad (\star)$$

where the coordinates both live in $\mathbb{Z}/t\mathbb{Z}$ and $\mathbb{Z}/p\mathbb{Z}$ respectively, with t a unit of time and p a unit of pitch. Thus, a score is defined as a finite, non-ordered set of distinct musical bars, and this description allows us to associate a point cloud with a score, where each point is then given by a musical bar containing the rhythmic and frequency information of the notes that are played.



(a) A sequence of Rips complexes for a point cloud data set representing an annulus. Upon increasing ϵ, holes appear and disappear. Which holes are real and which are noise?)

.



(b) [bottom] An example of the barcodes for $H_*(R)$ in the example of Figure 5a. [top] The rank of $H_k(R_{\epsilon_i})$ equals the number of intervals in the barcode for $H_k(R)$ intersecting the (dashed) line $\epsilon = \epsilon_i$.

Figure 5: Vietoris-Rips filtration and associated barcodes (extract from [27]).

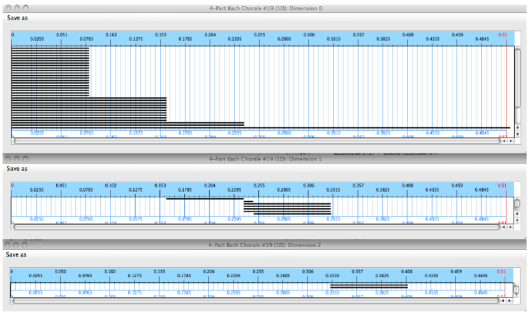


Figure 6: Barcodes for Bach's Chorale No. 19 (extract from [47]).

★ The automatic style classification problem

Among the major applications of topological analysis of musical data, we often find the question of automatic style classification, which will be the subject of Chapter 7 of this thesis. Bergomi's article [10] is a precursor, and we can mention as an example Figure 7 on which we observe a dendrogram obtained using persistence diagrams equipped with Bottleneck distance.

An important question here is about the metric on the set of barcodes: indeed, if the latter enable us to associate a topological signature with a piece of music, it seems necessary to define a distance on the set of these objects if the aim is to compare several pieces with each other. As we briefly stated above, the most popular metric on persistence diagrams is given by the Bottleneck distance (see for example [10], [19], or [32]). If BC_1 and BC_2 are two persistent diagrams (or barcodes), let $\phi : \mathrm{BC}_1 \to \mathrm{BC}_2$ be one of the applications which associates with each element of

BC_1 either an element of BC_2 if the barcodes have the same cardinalities, or its projection on the diagonal $x = y$ otherwise. The Bottleneck distance between BC_1 and BC_2 is then given by $(\star\star)$:

$$d_q(\mathrm{BC}_1, \mathrm{BC}_2) = \inf_{\phi \,:\, \mathrm{BC}_1 \to \mathrm{BC}_2} \left(\sup_{(b,d)\in \mathrm{BC}_1} \| \, (b,d) - \phi(b,d) \, \|_q \right) \qquad (\star\star)$$

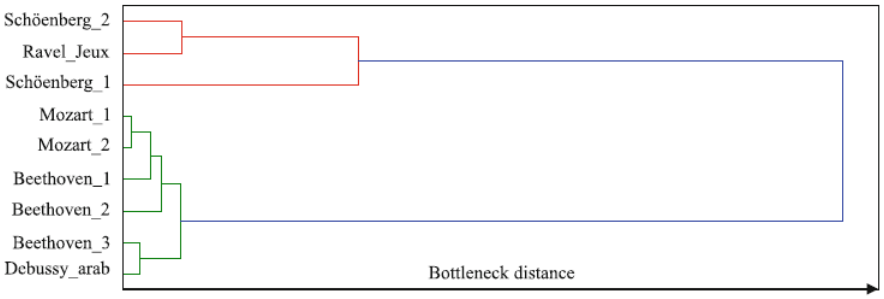


Figure 7: Persistence-based clustering of nine classical and modern pieces (extract from [10]).

Figure 8 from [32] shows an example of Bottleneck distance calculation. In this illustration, it is clear that the definition $(\star\star)$ comes from a geometric interpretation of persistence diagrams, but which leaves out the interpretation of barcodes from the point of view of homology class lifetimes and therefore bar lengths. In this thesis, our representation of persistent homology will be given exclusively by barcodes, which indeed seem to us closer to the interpretation in terms of class persistence, and indeed of bar length. As a result, the Bottleneck distance does not appear to be an intuitive tool for comparing such graphs, and we thus propose to compare two barcodes by computing statistics (mean, standard deviation and entropy) on these sets, as presented in the article [39]. This approach will enable us to produce satisfactory classification results, which will be presented in Chapter 7 of this manuscript.

It should also be noted that the problem of automatic classification of style and distance between barcodes is not the only subject to which persistent homology lends itself in the context of music analysis, and that it constitutes only one of the four applications proposed in this thesis. For example, the aim of Sethares was essentially to use barcodes to recover musical structures in sometimes artificially constructed pieces (such as scales), which is similar to the problems we focus on in Chapters 5 and 6 of this manuscript, in which our aim is to analyze barcodes from scores built on scales, Tonnetze or harmonized pieces extracted from Pop music.

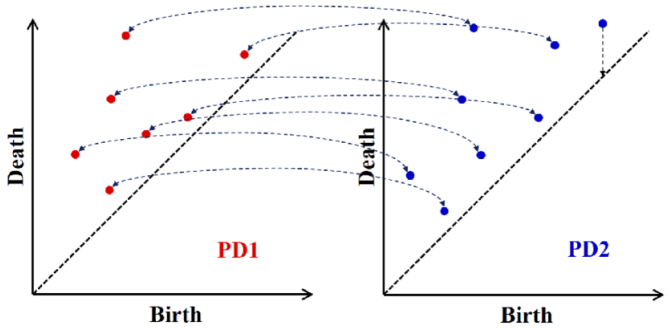


Figure 8: A schematic description how to calculate the bottleneck distance between two persistence diagrams (PDs). A point without a counterpart in the other PD is regarded as corresponding to the diagonal line (extract from [32]).

✯ From data anlysis to Discrete Fourier Transform

Associating a filtered complex with a piece of music requires extracting a point cloud, i.e. a collection of points in a metric space. As mentioned in the previous paragraphs, we will apply the Vietoris-Rips method to a point cloud built from the musical bars of a score. All that remains is to establish a distance between our vertices, and this is where the Discrete Fourier Transform (DFT) comes in.

The DFT is a mathematical tool for modeling simple musical sequences, such as scales or collections of rhythms. Historically, it was introduced into the context of music analysis by D. Lewin in 1959 (see [35]), whose aim was to formalize the notion of interval content between two sets of pitch-classes. This formalization leads to Lewin's Lemma, which we generalize in this thesis to the two-dimensional case (see Theorem 2.2.4). In the early 2000s, a series of works followed Lewin's last paper on the subject ([36]), including problems linked to the notion of homometry (see [37], [38]) or rhythmic canons (see [3], [55]). Following this, the DFT and its applications to music analysis became truly popular, notably thanks to the work of E. Amiot and J. Yust (see for example [4], [5], [54] and [56]).

In this thesis, we are primarily interested in modeling a musical sequence using the DFT: if $\mathcal{P} = (p_1, \dots, p_N)$ is a sequence of pitch-classes (or rhythmic data) in $\mathbb{Z}/n\mathbb{Z}$, then it can be associated with a list of n Fourier coefficients $\big(\mathcal{F}_\mathcal{P}(0), \dots, \mathcal{F}_\mathcal{P}(n-1)\big)$, where $\mathcal{F}_\mathcal{P}(x)$ is defined by the sum of complex exponentials

$$\mathcal{F}_\mathcal{P}(x) = \sum_{k \in \mathcal{P}} \exp\left(\frac{-2i\pi kx}{n}\right).$$

Thus, to compare two given musical sequences is to compare their respective lists of Fourier coefficients. Figure 9 from Amiot's book [4] illustrates the comparison between a melody and its accompaniment in a piece by Bartok. The musical sequences are modeled using the module of the Fourier coefficients.

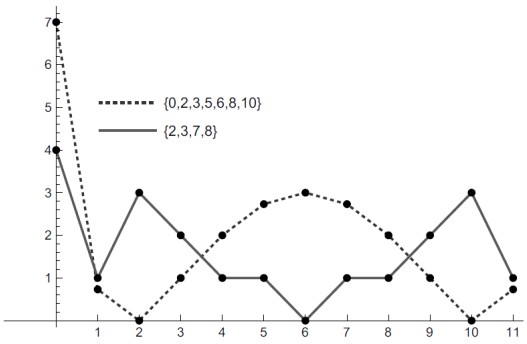


Figure 9: DFT magnitudes of melody and accompaniment in Bartok (extract from [4]).

As briefly mentioned above, our aim in this thesis is to model scores using a description given by their musical bars. In this context, the DFT seems promising because, by generalizing the previous model to the case of dimension two, we obtain a representation similar to the notion of musical bar we presented with $(\star)$, and comparing two such musical bars then amounts to comparing the associated Fourier coefficient matrices. Our model is therefore as follows: a musical score is a finite non-ordered set of distinct musical bars $\{\mathcal{B}_1, \dots \mathcal{B}_N\}$, and each musical bar is a subset of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, where t is a time-unit (for instance $t = 8$ if we choose the eighth note as unit) and p a pitch-unit (for instance $p = 12$ if we work in an octave). Let $\mathfrak{B}_{t,p}$ be the set of all musical bars in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. If $\mathcal{B}$ and $\mathcal{B}'$ are two elements of $\mathfrak{B}_{t,p}$, then their distance is given by

$$\mathrm{d}_{\mathrm{DFT}}(\mathcal{B}, \mathcal{B}') = \sum_{x=1}^{t} \sum_{y=1}^{p} |\mathcal{F}_\mathcal{B}(x,y) - \mathcal{F}_{\mathcal{B}'}(x,y)|$$

where $\mathcal{F}_{\mathcal{B}}(x,y)$ is the (x,y)-Fourier coefficient of $\mathcal{B}$ defined by

$$\mathcal{F}_{\mathcal{B}}(x,y) = \sum_{(k,l)\in\mathcal{B}} \exp\left(\frac{-2i\pi kx}{t}\right)\exp\left(\frac{-2i\pi ly}{p}\right).$$

Thus, $(\mathfrak{B}_{t,p}, \mathrm{d}_{\mathrm{DFT}})$ is a metric space, and a musical score becomes a point cloud, where each point is given by a subset of $(\mathfrak{B}_{t,p}, \mathrm{d}_{\mathrm{DFT}})$. Then, thanks to the Vietoris-Rips method presented above, we associate a complex filtration with a piece of music by means of the point cloud thus constructed, which then allows us to compute the persistent homology and extract a family of barcodes. This is the point of view that will be studied throughout this manuscript. On the other hand, this modeling offers the possibility of generalizing classical theoretical results on the DFT in dimension two, such as Lewin's lemma on interval content that we have already mentioned, or also Babbitt's Hexachord theorem ([6]).

Document layout

The manuscript is divided into three main parts: the first two are devoted to the study of the mathematical objects we used, namely persistent homology and the discrete Fourier transform, while the third is devoted to various musical applications.

In the first Part (I), we generalize the notion of DFT introduced by Lewin to the case of dimension two. More specifically, in Chapter 1, we make explicit the transition from a musical bar to a subset of $\mathbb{Z}/t\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}$, then define the notion of DFT-distance on the set of bars. In Chapter 2, we propose to generalize some classical theoretical results on the DFT such as Lewin's lemma on the interval content (Theorem 2.2.4) or Babbitt's Hexachord theorem (Theorem 2.3.1). Finally, we close this first part with a study of isometries on the set $(\mathfrak{B}_{t,p}, \mathrm{d}_{\mathrm{DFT}})$ of musical bars in $\mathbb{Z}/t\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}$ (Section 2.4).

The second part of this thesis (II) is divided into two chapters: in the first one (3), we give a few reminders of simplicial theory, then focus on the definition of filtration, persistent homology and, more specifically, barcodes. In the next Chapter (4), we define our model, that means extract a point cloud from a score by means of its musical bars and the DFT-distance.

In the third and final Part of this manuscript (III), we propose four musical applications of our model. The first is devoted to the inverse problem, that means applying our barcode extraction procedure to artificially constructed scores, such as scales or Tonnetze. Here, we find back some well-known topological features such as the torus on Euler's Tonnetz (Chapter 5). In Chapter 6, we propose to study harmonized pieces of Pop music by reducing them to just two tracks: the melody and accompaniment. This process makes it possible to associate a "graph-type" and a "complexity" with a given song. The third application is an essential part of this thesis, as it deals with the problem of automatic classification of musical style by means of persistent homology (Chapter 7), where our DFT-approach seems to provide some promising results. Finally, Chapter 8 is devoted to a different way of encoding of musical bars with another distance given by the Hausdorff metric, which notably provides satisfactory results on the global structure of certain pieces of music.

We end this manuscript with a general Conclusion and Appendices devoted to the codes used in our work (Appendice A.) and a presentation of our MIDI database (Appendice B.).

PART I

The two-dimensional Discrete Fourier Transform

Chapter 1.

The DFT for modeling basic musical structures

1.1. Introduction: the model

The Discrete Fourier Transform is commonly used in music theory as a model for basic musical structures such as scales, chords, rhythms, etc. The idea is the following: for a given musical sentence, let us say a musical bar, we can consider the set of **attack-times** or the **onsets** (horizontal row) or the set of **pitch-classes** (vertical column). In both cases, we need to choose a unit of time and a number of octaves to work with. For instance, in Figure 1.1.1, we consider a musical bar for which we can look at the set of attack-times as a subset of $\mathbb{Z}/8\mathbb{Z}$, if we take the quaver note as the unit of time (♪ = 1), or the set of pitch-classes as a subset of $\mathbb{Z}/12\mathbb{Z}$, if we choose to work in one octave (twelve notes).

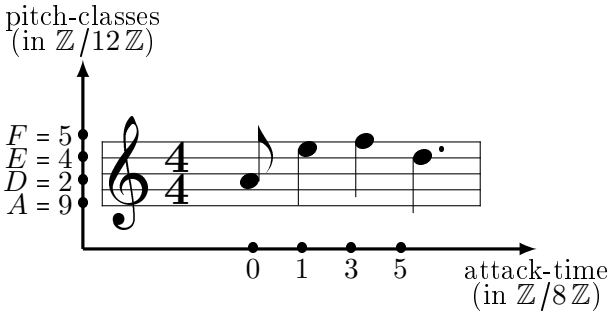


Figure 1.1.1: A musical bar where the x-axis and y-axis represent time and pitches, respectively. Using the quaver note as the time-unit (♪ = 1) and the octave as the ambitus (twelve notes), $\mathcal{T} = \{0, 1, 3, 5\}$ becomes the set of **attack-times** and $\mathcal{P} = \{9, 4, 5, 2\}$ becomes the set of **pitch-classes**.

Either by looking at the attack-times or the pitch-classes set, we can model these musical informations with functions of $\{0,1\}^{\mathbb{Z}/n\mathbb{Z}}$ (Definition 1.1.1) with the aim of computing the associated Discrete Fourier Transform (Definition 1.1.2).

Definition 1.1.1. Let $\mathcal{M} \subset \mathbb{Z}/n\mathbb{Z}$ be a musical structure (chords, scales, rhythms, etc.). The associated **characteristic map** $\mathbb{1}_{\mathcal{M}} : \mathbb{Z}/n\mathbb{Z} \to \{0,1\}$ is given by:

$$\mathbb{1}_{\mathcal{M}} : x \mapsto \begin{cases} 1 & \text{if } x \in \mathcal{M} \\ 0 & \text{otherwise.} \end{cases}$$

In general, a function of $A^{\mathbb{Z}/n\mathbb{Z}}$ for any subset $A \subset \mathbb{C}$ can be seen as the n-tuple of its value in each element of $\{0, 1, \ldots, n-1\}$. For instance, the associated characteristics maps with the

musical bar of Figure 1.1.1 (for attack-times set $\mathcal{T}$ and pitch-classes set $\mathcal{P}$) are respectively given by:

$$\mathbb{1}_{\mathcal{T}} : \{0,1,3,5\} \mapsto (1,1,0,1,0,1,0,0)$$

and

$$\mathbb{1}_{\mathcal{P}} : \{9,4,5,2\} \mapsto (0,0,1,0,1,1,0,0,0,1,0,0).$$

We can thus compute their associated Discrete Fourier Transform by following the next definition.

Definition 1.1.2. Let $\mathcal{M} \subset \mathbb{Z}/n\mathbb{Z}$ be a musical structure (chords, scales, rhythms, etc.).

1. The associated **Discrete Fourier Transform** $\mathcal{F}_{\mathcal{M}} : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$ (DFT) is the DFT of its characteristic map $\mathbb{1}_{\mathcal{M}} : \mathbb{Z}/n\mathbb{Z} \to \{0,1\}$ with the usual definition, that is:

$$\begin{array}{rccl} \mathcal{F}_{\mathcal{M}} = \widehat{\mathbb{1}_{\mathcal{M}}} & : & \mathbb{Z}/n\mathbb{Z} & \longrightarrow & \mathbb{C} \\ & & x & \longmapsto & \displaystyle\sum_{k=0}^{n-1} \mathbb{1}_{\mathcal{M}}(k) \exp\left(\tfrac{-2i\pi kx}{n}\right) = \sum_{k\in\mathcal{M}} \exp\left(\tfrac{-2i\pi kx}{n}\right) \end{array}$$

2. The associated **Fourier coefficients** are given by the n-tuple of the DFT values on $\mathbb{Z}/n\mathbb{Z}$:

$$\Big(\mathcal{F}_{\mathcal{M}}(0), \mathcal{F}_{\mathcal{M}}(1), \ldots, \mathcal{F}_{\mathcal{M}}(n-1)\Big)$$

With this definition, the DFTs associated with the musical bar of Figure 1.1.1 (for attack-times set $\mathcal{T}$ and pitch-classes set $\mathcal{P}$) are given by the following applications:

$$\begin{array}{rccl} \mathcal{F}_{\mathcal{T}} = \widehat{\mathbb{1}_{\mathcal{T}}} & : & \mathbb{Z}/8\mathbb{Z} & \longrightarrow & \mathbb{C} \\ & & x & \longmapsto & \displaystyle\sum_{k\in\mathcal{T}} \exp\left(\tfrac{-2i\pi kx}{8}\right) \end{array} \qquad \begin{array}{rccl} \mathcal{F}_{\mathcal{P}} = \widehat{\mathbb{1}_{\mathcal{P}}} & : & \mathbb{Z}/12\mathbb{Z} & \longrightarrow & \mathbb{C} \\ & & x & \longmapsto & \displaystyle\sum_{k\in\mathcal{P}} \exp\left(\tfrac{-2i\pi kx}{12}\right) \end{array}$$

In general, we are interested in the corresponding Fourier coefficients, which in some cases may be indicative of some musical information for basic structures such as chords, scales or rhythms (see [4] for more details in the topic). In our case, we will mostly use the Fourier coefficients as a metric to compare several musical bars from the same score together (see Section 4.2). In the example of Figure 1.1.1, the associated Fourier coefficients of attack-times $\mathcal{T} \subset \mathbb{Z}/8\mathbb{Z}$ and pitch-classes set $\mathcal{P} \subset \mathbb{Z}/12\mathbb{Z}$ are given respectively by

$$\Big(\mathcal{F}_{\mathcal{T}}(0), \mathcal{F}_{\mathcal{T}}(1), \ldots, \mathcal{F}_{\mathcal{T}}(7)\Big) = \left(4, 1 - \left(\frac{1}{2}i + \frac{1}{2}\right)\sqrt{2}, \ldots, 1 + \left(\frac{1}{2}i - \frac{1}{2}\right)\sqrt{2}\right)$$

and

$$\Big(\mathcal{F}_{\mathcal{P}}(0), \mathcal{F}_{\mathcal{P}}(1), \ldots, \mathcal{F}_{\mathcal{P}}(11)\Big) = \left(4, \frac{(1-2\sqrt{3})}{2}i - \frac{\sqrt{3}}{2}, \ldots, \frac{(2\sqrt{3}-1)}{2}i - \frac{\sqrt{3}}{2}\right).$$

Remark 1.1.3. Basic musical structures are also usually represented as n-gons, where n is the number of notes they contain. For instance, the corresponding illustration for the musical bar from Figure 1.1.1 is given in Figure 1.1.2. This type of representation for chords, scales, rhythms, etc. is often used to give a geometric representation of some common musical transformations such as translations or inversions (see Section 2.4). Also notice that, while this is an intuitive representation, it is also limited by the musical structure we are studying, such as when we consider chords together with single notes. Furthermore, we can imagine that the polygon we draw is not convex if we respect the order of appearance of the notes (see Figure 1.1.3).

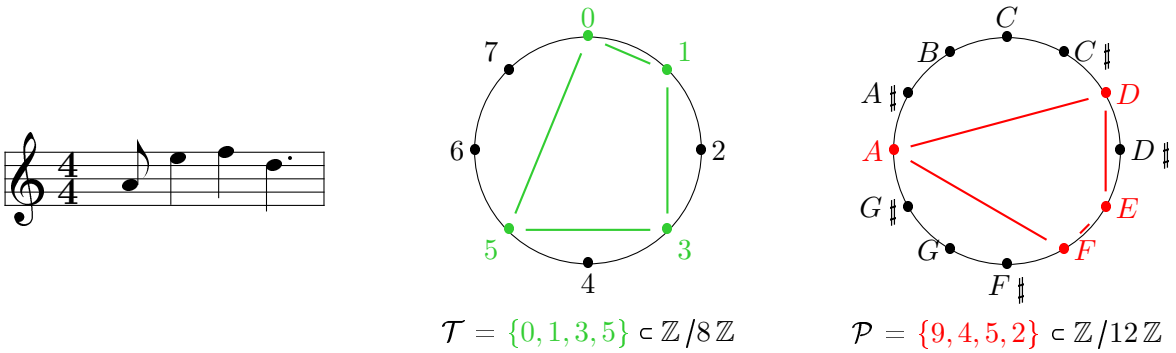


Figure 1.1.2: A musical bar and the representations of its attack-times set in $\mathbb{Z}/8\mathbb{Z}$ and its pitch-classes set in $\mathbb{Z}/12\mathbb{Z}$ as **convex** n-gons ($n = 4$).

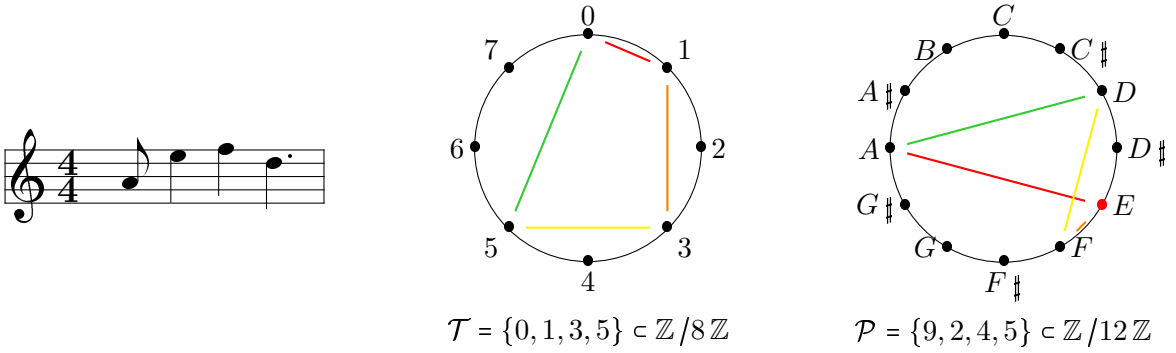


Figure 1.1.3: A musical bar and the representations its attack-times set in $\mathbb{Z}/8\mathbb{Z}$ and its pitch-classes set in $\mathbb{Z}/12\mathbb{Z}$ as **non-convex** n-gons ($n = 4$) by respecting the order in which the notes appear.

1.2. The two-dimensional DFT

The idea now is to use the previous model that is given by the DFT for attack-times set $\mathcal{T}$ and pitch-classes set $\mathcal{P}$ by combining both representations. In other words, we will consider $\mathcal{T}$ and $\mathcal{P}$ together as a subset of $\mathbb{Z}/8\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$, and we will call it a **musical bar**, denoted first by $\mathcal{B}_{\mathcal{T},\mathcal{P}}$. Let us take back our example of Figure 1.1.1: Figure 1.2.1 gives a representation of attack-times set $\mathcal{T} = \{0,1,3,5\}$ and pitch-classes set $\mathcal{P} = \{9,4,5,2\}$ together as a subset of $\mathbb{Z}/8\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$ with the following description:

$$\mathcal{B}_{\mathcal{T},\mathcal{P}} = \{(0,9),(1,4),(3,5),(5,2)\} \subset \mathbb{Z}/8\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$$

This two-dimensional representation leads to the natural definition of a **musical bar**, which is the musical structure we will always consider in the persistent homology part (see Section II). For this purpose, we need to fix a **unit of time** t, which can be chosen by taking the shortest note in the bar (sometimes called the **elementary rhythm**) and respecting the correspondence of Table 1.2.1.

𝅝	𝅗𝅥	♩	♪	𝅘𝅥𝅯	𝅘𝅥𝅰
1	2	4	8	16	32

Table 1.2.1: The table of correspondence between basic note durations and time-units.

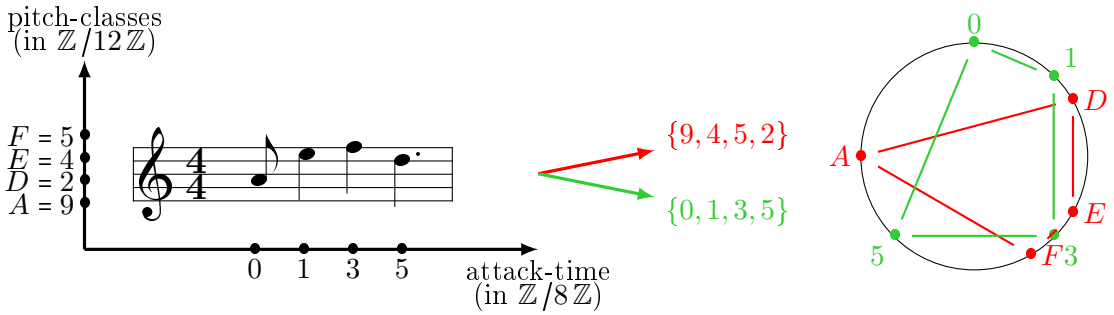
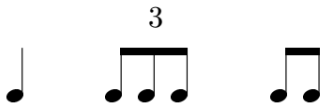


Figure 1.2.1: A musical sentence bar with its associated combined two-dimensional modeling $B_{\mathcal{T},\mathcal{P}} = \{(0,9),(1,4),(3,5),(5,2)\}$ in $\mathbb{Z}/8\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$

In this table, we give only the usual rhythms, but we can also choose a particular time-unit when there are irregular divisions like tuplets: for instance, if we want to encode the following sequence of rhythms



we need to consider that the quarter note ♩ is divided in half and third, so the sentence is given by the following beats sequence

$$1 \quad \frac{1}{3}\ \frac{1}{3}\ \frac{1}{3} \quad \frac{1}{2}\ \frac{1}{2}.$$

Thus, by taking the least common multiple of the denominator, we get our unit of time, which here is $t = 2 \times 3$ and in this example $\mathcal{T} = \{0,6,8,10,12,15,18\}$. This method works for any irregular subdivision of time in a musical bar. We also need to fix th number of notes p we are working with. Instead of taking one single octave and setting $\mathcal{P}$ in $\mathbb{Z}/12\mathbb{Z}$, as is often done in the one-dimensional case, we will rather consider the number of octaves $m \in \mathbb{N} \setminus \{0\}$ in which all the notes live and therefore set $p = 12m$. We will call this number p the **ambitus**.

Remark 1.2.1. Notice that, in this introduction, we rather speak about $\mathcal{T}$ and $\mathcal{P}$ as sets from which we build a musical bar $\mathcal{B}_{\mathcal{T},\mathcal{P}}$, but in practice there is no canonical way to reconstruct $\mathcal{T}$ and $\mathcal{P}$ from this musical bar. For instance, if $\mathcal{B}_{\mathcal{T},\mathcal{P}}$ consists of only one chord with onset 0, such as $\mathcal{B}_{\mathcal{T},\mathcal{P}} = \{(0,0),(0,4),(0,7)\}$ for example, then $\mathcal{T}$ should only contain the onset 0 but with multiplicity 3. There is also the problem of order, since we need to associate the correct elements of $\mathcal{T}$ with those of $\mathcal{P}$. If it is possible to find $\mathcal{T}$ and $\mathcal{P}$ from $\mathcal{B}_{\mathcal{T},\mathcal{P}}$ in the simplest cases, as in Figure 1.2.1, where all the onsets and pitches are different, in practice we would not need to work with $\mathcal{T}$ and $\mathcal{P}$. Indeed, we will only be focusing on a musical bar as a subset of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ in itself, and so we will simply denote this object by $\mathcal{B}$ and forget about $\mathcal{T}$ and $\mathcal{P}$.

We can now give the proper definition of a musical bar.

Definitions 1.2.2. Let t be a **unit of time** and p an **ambitus**, both fixed.

1. A **musical bar** $\mathcal{B}$ with n elements is a subset of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ given by

$$\mathcal{B} = \{(t_1,p_1),\dots,(t_n,p_n)\}$$

where an element (t_j,p_j) of $\mathcal{B}$ is called a **note** which is characterized by two coordinates:

i) its **attack-time** or **onset** t_j modulo t

ii) its **pitch-class** p_j modulo p

2. A **musical score** $\mathcal{S}$ is the non-ordered set of its N distinct musical bars modulo (t,p):

$$\mathcal{S} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_N\} \text{ with } \mathcal{B}_i \subset \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \text{ and } \mathcal{B}_i \neq \mathcal{B}_j \text{ if } i \neq j$$

We will denoted by $\mathfrak{B}_{t,p}$ the set of all the musical bars in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.

Remark 1.2.3. We can also describe any musical bar by using a polynomial point of view: let $\mathcal{B}$ be a musical bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. The **characteristic polynomial** associated with $\mathcal{B}$ is given by

$$P_{\mathcal{B}}(X,Y) = \sum_{(k,l)\in\mathcal{B}} X^k Y^l \quad \in \quad \mathbb{C}[X,Y]\Big/(X^t-1, Y^p-1)$$

Basically, attack-times are encoded by the power of the first variable X while pitch-classes are represented by the power of the second variable Y. For instance, the associated characteristic polynomial with the musical bar $\mathcal{B}$ from Figure 1.2.1 is given by

$$P_{\mathcal{B}}(X,Y) = Y^9 + XY^4 + X^2Y^5 + X^3 \quad \in \quad \mathbb{C}[X,Y]\Big/(X^8-1, Y^{12}-1)$$

This idea was already developed in [4] for basic musical sentences in $\mathbb{Z}/n\mathbb{Z}$ in the one-dimensional case.

Remark 1.2.4. Definition 1.2.2 is arbitrary but based on the natural idea of generalizing a theory that already exists in the one-dimensional case, where we study rhythms and pitches separately by working respectively with attack-times and pitch-classes sets (see [4]). However, our definition does not provide a bijection between subsets of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ and $\mathfrak{B}_{t,p}$: in fact, if we can describe any musical bar of a given score by characterizing a note with its position and its pitch in the bar, there are several musical bars that will correspond to the obtained description. To get the bijection, one can add for example the duration of the note as a third coordinate, as we will make in the Hausdorff part (see Section 8).

We can now generalize the notion of DFT from one-dimension in two and get a new description of any musical bar of a given score using a matrix of Fourier coefficients.

Definitions 1.2.5. Let $B = (\mathcal{T}, \mathcal{P})$ be a musical bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.

1. The associated **characteristic map** $\mathbb{1}_{\mathcal{B}} : \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ is given by:

$$\mathbb{1}_{\mathcal{B}} : (x,y) \mapsto \begin{cases} 1 & \text{if } (x,y) \in \mathcal{B} \\ 0 & \text{otherwise.} \end{cases}$$

2. The associated **DFT** is the DFT of the characteristic map $\mathbb{1}_{\mathcal{B}}$:

$$\begin{array}{rccl} \mathcal{F}_{\mathcal{B}} = \widehat{\mathbb{1}_{\mathcal{B}}} & : & \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} & \longrightarrow \quad \mathbb{C} \\ & & (x,y) & \longmapsto \quad \displaystyle\sum_{(k,l)\in\mathcal{B}} \exp\left(\tfrac{-2i\pi kx}{t}\right) \exp\left(\tfrac{-2i\pi ly}{p}\right) \end{array}$$

3. The associated **Fourier coefficients** are given by the following matrix $M_{\mathcal{B}} \in \mathcal{M}_{t\times p}(\mathbb{C})$:

$$M_{\mathcal{B}} = \Big(\mathcal{F}_{\mathcal{B}}(x,y)\Big)_{(x,y)\in\mathbb{Z}/t\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}}$$

Remark 1.2.6. The DFT of a musical bar $\mathcal{B}$ can also be alternatively defined by its **real** and **imaginary** values: for every pair $(x,y) \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we have

$$\begin{aligned} \Re\big(\mathcal{F}_{\mathcal{B}}(x,y)\big) &= \sum_{(k,l)\in\mathcal{B}} \cos\left(\frac{2\pi kx}{t} + \frac{2\pi ly}{p}\right) \\ \Im\big(\mathcal{F}_{\mathcal{B}}(x,y)\big) &= -\sum_{(k,l)\in\mathcal{B}} \sin\left(\frac{2\pi kx}{t} + \frac{2\pi ly}{p}\right) \end{aligned}$$

and we have the **magnitude** of a musical bar $\mathcal{B}$:

$$|\mathcal{F}_{\mathcal{B}}| = \sqrt{\Re\big(\mathcal{F}_{\mathcal{B}}\big)^2 + \Im\big(\mathcal{F}_{\mathcal{B}}\big)^2}$$

Remark 1.2.7. This remark follows the 1.2.3 one: recall that we can associate a characteristic polynomial with any musical bar $\mathcal{B}$. From this polynomial $P_{\mathcal{B}}$, we can also find back the Fourier coefficients associated with $\mathcal{B}$ using the following map:

$$\begin{array}{ccc} \mathbb{C}[X,Y]/(X^t-1, Y^p-1) & \to & \mathbb{C}^{tp} \\ P_{\mathcal{B}} & \mapsto & \mathcal{F}_{\mathcal{B}} \end{array}$$

and taking the polynomial evaluation for every pair $(k,l) \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$:

$$\mathcal{F}_{\mathcal{B}}(k,l) = P_{\mathcal{B}}\big(e^{\frac{-2ik\pi}{t}}, e^{\frac{-2il\pi}{p}}\big)$$

In this chapter, we defined the notion of a musical bar $\mathcal{B}$ as a subset of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, and it was associated with a matrix of Fourier coefficients $M_{\mathcal{B}}$. In the musical applications, we will use this matrix to compare several musical bars together: indeed, we can also define a metric on the set $\mathfrak{B}_{t,p}$ using the DFT, and this metric will be the basis of the computations for the next parts of this manuscript. We define this metric now to prove some theoretical results for the next chapter.

Definition 1.2.8. Let $\mathcal{B}$ and $\mathcal{B}'$ be two elements of $\mathfrak{B}_{t,p}$. The **DFT-distance** between $\mathcal{B}$ and $\mathcal{B}'$ is given by the 1-distance between their respective Fourier coefficients matrices:

$$\mathrm{d}_{\mathrm{DFT}}(\mathcal{B},\mathcal{B}') = \| M_{\mathcal{B}} - M_{\mathcal{B}'} \|_1 = \sum_{x=1}^{t}\sum_{y=1}^{p} |\mathcal{F}_{\mathcal{B}}(x,y) - \mathcal{F}_{\mathcal{B}'}(x,y)|$$

We denoted by $(\mathfrak{B}_{t,p}, \mathrm{d}_{\mathrm{DFT}})$ the metric space of all the musical bars of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ equipped with the DFT-distance.

We are now able to model any musical bar as a subset of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ and use the DFT to associate a matrix of Fourier coefficients with it. We have also defined a metric for comparing two musical bars together. Starting from this, we have two general lines of research: a natural thing to do is to try to generalize some theoretical results from the one-dimensional case (see next Chapter 2) or we can use this new approach in the context of persistent homology and Topological Data Analysis by using this new metric from 1.2.8 in order to extract a point cloud from a given score (see Chapter 4 and Part III).

Chapter 2.

Generalization of theoretical results

In this chapter, our purpose is to generalize some musical properties that the DFT already satisfies in one dimension, such as the interval content our the Hexachordal Theorem. In fact, most of the properties that are true in one dimension are also satisfied in higher dimensions, which is treated in [50] (Generalization of the DFT in the case of abelian groups): indeed, all the basic theorems can be extended to this case, and here we simply recall some basic facts (**inversion** and **convolution**) that we need for our musical applications. Notice that we are doing this for any application $f : \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$, and we will use it in the case of the characteristic map associated with a musical bar $\mathcal{B} \in Ztp$.

Theorem 2.0.1 (DFT-inversion in two dimensions)**.** *Let f be an application from $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ to $\mathbb{C}$. The DFT of f is given by the application*

$$\begin{array}{rccl} \mathcal{F}(f) = \widehat{f} \ : & \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} & \longrightarrow & \mathbb{C} \\ & (x,y) & \longmapsto & \sum_{k=0}^{t-1} \sum_{l=0}^{p-1} f(x,y) \exp\left(\frac{-2i\pi kx}{t}\right) \exp\left(\frac{-2i\pi ly}{p}\right) \end{array}$$

Then, the two-dimensional DFT is a linear automorphism of $\mathbb{C}^{tp}$, and its reciprocal map $\mathcal{F}^{-1}$ is given by

$$\begin{array}{rccl} \mathcal{F}^{-1}(\widehat{f}) \ : & \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} & \longrightarrow & \mathbb{C} \\ & (x,y) & \longmapsto & \frac{1}{tp}\left(\sum_{k=0}^{t-1} \sum_{l=0}^{p-1} \widehat{f}(x,y) \exp\left(\frac{2i\pi kx}{t}\right) \exp\left(\frac{2i\pi ly}{p}\right)\right) \end{array}$$

Theorem 2.0.2 (DFT-convolution in two dimensions)**.** *Let f and g be two applications from $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ to $\mathbb{C}$. We define the convolution product of f and g by*

$$\begin{array}{rccl} f \star g \ : & \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} & \longrightarrow & \mathbb{C} \\ & (x,y) & \longmapsto & \sum_{k=0}^{t-1} \sum_{l=0}^{p-1} f(x-k, y-l) g(k,l) \end{array}$$

Then, the DFT of a convolution product is the termwise product of the DFTs of the maps:

$$\forall (x,y) \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}, \quad \widehat{f \star g}(x,y) = \widehat{f}(x,y) \times \widehat{g}(x,y)$$

Remark 2.0.3. Recall Remarks 1.2.3 and 1.2.7: any musical bar $\mathcal{B}$ has an associated characteristic polynomial $P_{\mathcal{B}}$. Now if $\mathcal{B}$ and $\mathcal{B}'$ are two musical bars, then the convolution product of $\mathcal{B}$ and $\mathcal{B}'$ corresponds to the convolution product defined in 2.0.2 for the characteristic maps $\mathbb{1}_{\mathcal{B}}$ and $\mathbb{1}_{\mathcal{B}'}$:

$$\mathcal{B} \star \mathcal{B}' := \mathbb{1}_{\mathcal{B}} \times \mathbb{1}_{\mathcal{B}'}$$

Then, this convolution product can alternatively be defined by the termwise product of the characteristic polynomial:

$$\mathcal{B} \star \mathcal{B}' := P_{\mathcal{B}} \times P_{\mathcal{B}'}$$

2.1. Basic properties

This first section generalizes some basic properties that the DFT satisfies in the one-dimensional case. It corresponds to the first chapter of [4] and it will be useful for the next sections of this chapter. We will use it to generalize Lewins's Lemma and Hexachordal Theorem.

Proposition 2.1.1. *Let $\mathcal{B}$ be a musical bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.*

1. $\mathcal{F}_{\mathbb{Z}/t\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}}(0,0) = tp$ *and* $\mathcal{F}_{\mathcal{B}}(0,0) = \mathrm{card}(\mathcal{B})$.
2. *The DFT of a musical bar with only one element $\mathcal{B} = \{(a,b)\}$ is given by the product of complex exponential:*
$$\mathcal{F}_{\mathcal{B}}(x,y) = \exp\left(\frac{-2i\pi ax}{t}\right)\exp\left(\frac{-2i\pi by}{p}\right)$$
3. *The DFT of a musical bar containing all possible notes, i.e.*
$$\mathcal{B} = \{(a,b) \mid (a,b) \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}\}$$
is given by
$$\mathcal{F}_{\mathcal{B}}(x,y) = \begin{cases} 0 & \text{if } (x,y) \neq (0,0) \\ tp & \text{else} \end{cases}$$

Proof. 1. This comes directly from the definition of the DFT.

2. This is a sum with only one element.

3. The coefficient for $(x,y) = (0,0)$ is given by the first point. For $(x,y) \neq (0,0)$, we have

$$\begin{aligned} \mathcal{F}_{\mathcal{B}}(x,y) &= \sum_{(k,l)\in\mathcal{B}} \exp\left(\tfrac{-2i\pi kx}{t}\right)\exp\left(\tfrac{-2i\pi ly}{p}\right) \\ &= \sum_{k=0}^{t-1}\sum_{l=0}^{p-1} \exp\left(\tfrac{-2i\pi kx}{t}\right)\exp\left(\tfrac{-2i\pi ly}{p}\right) \\ &= \sum_{k=0}^{t-1} \exp\left(\tfrac{-2i\pi kx}{t}\right)\sum_{l=0}^{p-1}\exp\left(\tfrac{-2i\pi ly}{p}\right) \\ &= 0 \end{aligned}$$

■

In order to generalize Hexachordal Theorem in the two-dimensional case, we need to define the complement of a musical bar.

Definition 2.1.2. Let $\mathcal{B} \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ be a musical bar. We define the **complement musical bar** $\overline{\mathcal{B}}$ of $\mathcal{B}$ as follow:
$$\overline{\mathcal{B}} = \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \setminus \mathcal{B}$$

Example 2.1.3. 1. For $t = 1$ and $p = 12$, $\mathcal{B}$ is a *chord bar*, *i.e.* a musical bar containing only one group of n notes at the same position 0. For instance, if $\mathcal{B}$ is the chord corresponding to the C-Major scale (the white keys on a piano),
$$\mathcal{B} = \{(0,0),(0,2),(0,4),(0,5),(0,7),(0,9),(0,11)\} \in \mathbb{Z}/1\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$$
then the complement $\overline{\mathcal{B}}$ is the chord that corresponds to the pentatonic scale (the black keys)
$$\overline{\mathcal{B}} = \{(0,1),(0,3),(0,6),(0,8),(0,10)\} \in \mathbb{Z}/1\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}.$$

2. For $t = 2$ and $p = 12$, $\mathcal{B}$ is a bar that might contain elements on position 0 and position 1 (two half notes 𝅗𝅥). For instance, we can build the musical bar $\mathcal{B}$ with white keys in position 0 and black keys in position 1:

$$\mathcal{B} = \left\{ \begin{array}{c} (0,0),(0,2),(0,4),(0,5),(0,7),(0,9),(0,11), \\ (1,1),(1,3),(1,6),(1,8),(1,10) \end{array} \right\} \in \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$$

Thus the complement bar $\overline{\mathcal{B}}$ is the musical bar with the same chords with reversed positions:

$$\overline{\mathcal{B}} = \left\{ \begin{array}{c} \{(0,1),(0,3),(0,6),(0,8),(0,10), \\ (1,0),(1,2),(1,4),(1,5),(1,7),(1,9),(1,11) \end{array} \right\} \in \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$$

In this particular case, $\text{card}(\mathcal{B}) = \text{card}(\overline{\mathcal{B}}) = \frac{2\times 12}{2} = 12$.

Proposition 2.1.4. *Let $\mathcal{B}$ be a musical bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. For all pairs $(x,y) \neq (0,0)$ in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we have*

$$\mathcal{F}_{\overline{\mathcal{B}}}(x,y) = -\mathcal{F}_{\mathcal{B}}(x,y)$$

Proof. For each pair $(x,y) \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we have the sum

$$\mathcal{F}_{\mathbb{Z}/t\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}}(x,y) = \mathcal{F}_{\mathcal{B}}(x,y) + \mathcal{F}_{\overline{\mathcal{B}}}(x,y)$$

and with Proposition 2.1.1, for $(x,y) \neq (0,0)$ we get that $\mathcal{F}_{\overline{\mathcal{B}}}(x,y) = -\mathcal{F}_{\mathcal{B}}(x,y)$. ∎

2.2. Intervallic content

We first recall the definition in the one-dimensional case.

Definition 2.2.1. Let $\mathcal{M} \subset \mathbb{Z}/n\mathbb{Z}$ be a musical structure (chords, scales, rhythms, etc.). The **interval content** associated with $\mathcal{M}$ is given by the following function of $\mathbb{N}^{\mathbb{Z}/n\mathbb{Z}}$:

$$\begin{array}{rccl} \text{IC}_{\mathcal{M}} & : & \mathbb{Z}/n\mathbb{Z} & \to & \mathbb{N} \\ & & k & \mapsto & \text{card}\{(i, i+k) \in \mathcal{M} \times \mathcal{M}\} \end{array}$$

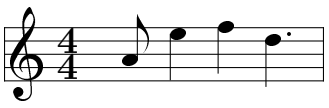
This definition can easily be adapted to the two-dimensional case:

Definition 2.2.2. Let $\mathcal{B}$ be a musical bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. The **interval content** associated with $\mathcal{B}$ is defined by:

$$\begin{array}{rccl} \text{IC}_{\mathcal{B}} & : & \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} & \longrightarrow & \mathbb{N} \\ & & (k,l) & \longmapsto & \#\left\{\left((i,j),(i+k,j+l)\right) \in \mathcal{B}^2\right\} \end{array}$$

In practice, for a pair $(k,l) \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, $\text{IC}(k,l)$ is the number of pairs $\left((i,j),(t,u)\right)$ that are shifted exactly by k on the time-axis modulo t and exactly by l in the pitch-axis modulo p.

Example 2.2.3. Recall the musical bar $\mathcal{B} = \{(0,9),(1,4),(3,5),(5,2)\}$ from Figure 1.1.1:



Its interval content (in $\mathbb{Z}/8\mathbb{Z}\times\mathbb{Z}/12\mathbb{Z}$) is given by the following 8×12 matrix:

$$\begin{pmatrix} 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

For instance, $\mathrm{IC}_{\mathcal{B}}(2,1) = 1$ and it corresponds to the pair $\big((1,4),(3,5)\big)$.

In one dimension, there is a musical interpretation of the interval content given by Lewin ([35]): basically, for one pitch-class set $\mathcal{P}$, $\mathrm{IC}_{\mathcal{P}}$ measures the probability of occurrence of an interval between two notes in $\mathcal{P}$. In our case, it does the same but it takes into account the onsets and the pitches at the same time. We can now give a generalization of the the Lewin's Lemma in this context:

Theorem 2.2.4 (Lewin's Lemma in two dimensions)**.** *Let $\mathcal{B}$ be a musical bar in $\mathbb{Z}/t\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}$. The DFT of the interval content of $\mathcal{B}$ is equal to the square of the magnitude of the DFT of $\mathcal{B}$:*

$$\widehat{\mathrm{IC}_{\mathcal{B}}} = |\mathcal{F}_{\mathcal{B}}|^2$$

Proof. Let $(k,l)\in\mathbb{Z}_t p$. Using the definition of a convolution product, we get that

$$\begin{aligned}
\mathrm{IC}_{\mathcal{B}}(k,l) &= \#\left\{\big((i,j),(i+k,j+l)\big)\in\mathcal{B}^2,\ (i,j)\in\mathcal{B}\right\} \\
&= \sum_{(i,j)\in\mathcal{B}} \mathbb{1}_{\mathcal{B}}(i,j)\,\mathbb{1}_{\mathcal{B}}(i+k,j+l) \\
&= \sum_{(i,j)\in\mathcal{B}} \mathbb{1}_{-\mathcal{B}}(-i,-j)\,\mathbb{1}_{\mathcal{B}}\big(k-(-i),l-(-j)\big) \\
&= \sum_{(t,u)\in-\mathcal{B}} \mathbb{1}_{-\mathcal{B}}(t,u)\,\mathbb{1}_{\mathcal{B}}(k-t,l-t) \\
&= \mathbb{1}_{-\mathcal{B}}\star\mathbb{1}_{\mathcal{B}}(k,l)
\end{aligned}$$

Then, by computing the DFT of IC and using Theorem 2.0.2, we get that

$$\begin{aligned}
\widehat{\mathrm{IC}_{\mathcal{B}}}(k,l) &= \widehat{\mathbb{1}_{-\mathcal{B}}\star\mathbb{1}_{\mathcal{B}}}(k,l) \\
&= \widehat{\mathbb{1}}_{-\mathcal{B}}(k,l)\times\widehat{\mathbb{1}}_{\mathcal{B}}(k,l) \\
&= \overline{\widehat{\mathbb{1}_{\mathcal{B}}}(k,l)}\times\widehat{\mathbb{1}_{\mathcal{B}}}(k,l) \\
&= |\widehat{\mathbb{1}}_{\mathcal{B}}(k,l)|^2 \\
&= |\mathcal{F}_{\mathcal{B}}(k,l)|^2
\end{aligned}$$

■

2.3. Hexachordal Theorem

In this section, we aim to generalize the proof of Babbitt's Hexachordal Theorem ([6]) in the two-dimensional case. In fact, the proof given by [4] essentially uses Lewin's lemma, which we have just generalized.

Theorem 2.3.1 (Hexachordal Theorem in two dimensions)**.** *Let $\mathcal{B}$ be a musical bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ and $\overline{\mathcal{B}}$ its complement bar. We assume that* $\mathrm{card}(\mathcal{B}) = \mathrm{card}(\overline{\mathcal{B}})$. *Then, $\mathcal{B}$ and $\overline{\mathcal{B}}$ have the same interval content:*

$$IC_{\mathcal{B}} = IC_{\overline{\mathcal{B}}}$$

Proof. By Proposition 2.1.1, we have

$$\mathcal{F}_{\mathcal{B}}(0,0) = \mathrm{card}(\mathcal{B}) = \mathrm{card}(\overline{\mathcal{B}}) = \mathcal{F}_{\overline{\mathcal{B}}}(0,0)$$

and Proposition 2.1.4 says that, for any $(x,y) \neq (0,0)$,

$$\mathcal{F}_{\overline{\mathcal{B}}}(x,y) = -\mathcal{F}_{\mathcal{B}}(x,y).$$

We thus have for any $(x,y) \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$

$$\left|\mathcal{F}_{\mathcal{B}}(x,y)\right| = \left|\mathcal{F}_{\overline{\mathcal{B}}}(x,y)\right|.$$

Therefore, by applying Lewin's Lemma in two dimensions 2.2.4, we get that

$$\widehat{\mathrm{IC}}_{\mathcal{B}} = \left|\mathcal{F}_{\mathcal{B}}\right|^2 = \left|\mathcal{F}_{\overline{\mathcal{B}}}\right|^2 = \widehat{\mathrm{IC}}_{\overline{\mathcal{B}}}$$

and by DFT-inverse from Theorem 2.0.1, $\mathrm{IC}_{\mathcal{B}} = \mathrm{IC}_{\overline{\mathcal{B}}}$. ∎

Example 2.3.2.

1. For $t = 1$ and $p = 12$, we find the classical Hexachordal Theorem in one dimension. For instance, if $\mathcal{B}$ is the musical bar containing only one chord with 6 elements

   $$\mathcal{B} = \{(0,0),(0,2),(0,3),(0,7),(0,8),(0,10)\}$$

   with its complement bar $\overline{\mathcal{B}}$

   $$\overline{\mathcal{B}} = \{(0,1),(0,4),(0,5),(0,6),(0,9),(0,11)\},$$

   then we have

   $$IC_{\mathcal{B}} = IC_{\overline{\mathcal{B}}} = (6,2,3,2,3,4,2,4,3,2,3,2).$$

   This interval content is the same as the one associated with the pitch-classes set $(2,3,7,8,10)$ and its complement, which are both hexachords (see Figure 2.3.1).

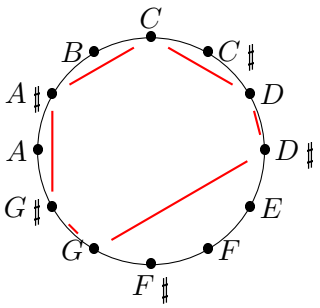
2. For $t = 2$ and $p = 12$, a hexachord contains 12 elements. For instance, the musical bar $\mathcal{B}$ from the second point of Example 2.1.3 has white keys at position 0 and black keys at position 1:

   $$\mathcal{B} = \left\{ \begin{array}{c} (0,0),(0,2),(0,4),(0,5),(0,7),(0,9),(0,11), \\ (1,1),(1,3),(1,6),(1,8),(1,10) \end{array} \right\}$$

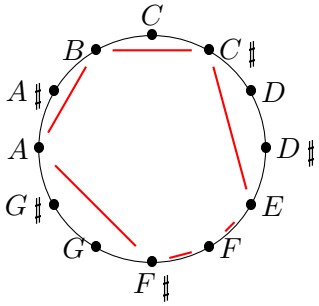
   We have $\mathrm{card}(\mathcal{B}) = 12$ and $\overline{\mathcal{B}}$ is the same bar with the chords position reverse. The corresponding interval content is given by

   $$IC_{\mathcal{B}} = IC_{\overline{\mathcal{B}}} = \begin{pmatrix} 12 & 2 & 8 & 6 & 4 & 10 & 2 & 10 & 4 & 6 & 8 & 2 \\ 0 & 10 & 4 & 6 & 8 & 2 & 10 & 2 & 8 & 6 & 4 & 10 \end{pmatrix}$$

3. In general, for any t and p, we can always construct a hexachord $\mathcal{B}$ containing $\frac{tp}{2}$ elements with a $\frac{p}{2}$-chord at each position of the bar, and the complement $\overline{\mathcal{B}}$ will have the same interval content.



$\mathcal{P} = \{0, 2, 3, 7, 8, 10\}$



$\overline{\mathcal{P}} = \{1, 4, 5, 6, 9, 11\}$

Figure 2.3.1: Two hexachords $\mathcal{B}$ and $\overline{\mathcal{B}}$ in $\mathbb{Z}/1\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$ reduces to their pitch-classes set in one dimension.

2.4. Musical bar transformations and isometries

In this section, we are interested in the basic transformations we can perform on a musical bar. More precisely, it seems natural to focus on which of them preserve the metric we defined with the DFT.

2.4.1. Rotations and reflections

Definitions 2.4.1. Let $\mathcal{B}$ be a musical bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ and (α, β) be an element of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.

i) A (α, β)-**rotation** of $\mathcal{B}$ is a map $\rho_{\alpha,\beta} : \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ such that, for all pairs (x, y) in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we have

$$\rho_{\alpha,\beta}(x, y) = (\alpha + x, \beta + y)$$

and thus

$$\rho_{\alpha,\beta} \cdot \mathcal{B} = \{(\alpha + t, \beta + p),\ (t, p) \in \mathcal{B}\}.$$

ii) A (α, β)-**reflection** of $\mathcal{B}$ is a map $\sigma_{\alpha,\beta} : \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ such that, for all pairs (x, y) in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we have

$$\sigma_{\alpha,\beta}(x, y) = (\alpha - x, \beta - y)$$

and thus

$$\sigma_{\alpha,\beta} \cdot \mathcal{B} = \{(\alpha - t, \beta - p),\ (t, p) \in \mathcal{B}\}.$$

Examples 2.4.2. In the following examples, we will consider the musical bar

$$\mathcal{B} = \{(0, 0), (1, 2), (2, 4), (3, 5)\}$$

with time-unit 4 (one quarter note ♩) and pitch-unit 12 (one octave). In other words, all rotations and reflections will go from $\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$ to itself.

1) For $\alpha = 1,\ \beta = 0$, $\rho_{1,0}$ is a $(1, 0)$-rotation of $\mathcal{B}$:

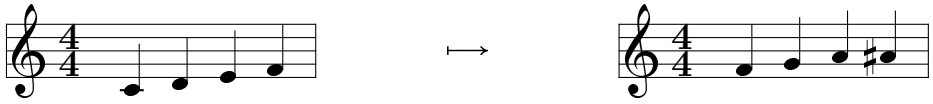
$$\rho_{1,0} \cdot \mathcal{B} = \{(0, 5), (1, 0), (2, 2), (3, 4)\}$$

which gives us the following permutation of the notes:

2) For $\alpha = 0,\ \beta = 5$, $\rho_{0,5}$ is a $(0,5)$-rotation of $\mathcal{B}$:

$$\rho_{0,5} \cdot \mathcal{B} = \{(0,5),(1,7),(2,9),(3,10)\}$$

which gives us the following transposition of the notes:



3) For $\alpha = 2,\ \beta = 2$, $\sigma_{2,2}$ is a $(2,2)$-reflection of $\mathcal{B}$:

$$\sigma_{2,2} \cdot \mathcal{B} = \{(2,2),(1,0),(0,10),(3,9)\}$$

which gives us the following transposition and permutation of the notes:

Remarks 2.4.3. 1) If $\beta = 0$, then $\rho_{\alpha,0}$ affects only the attack-times of the musical bar, while if $\alpha = 0$, then $\rho_{0,\beta}$ affects only the pitches. In the first case, we will call $\rho_{\alpha,0}$ a **permutation** of the musical bar (a cyclic permutation of the attack-times) and in the second case, we will say that $\rho_{0,\beta}$ is a **transposition** (a pitches transposition of the musical bar). Notice that a rotation can act on attack-times or pitch-classes separately, while a reflection always changes both sets.

2) Every reflection $\sigma_{\alpha,\beta}$ satisfies the relation

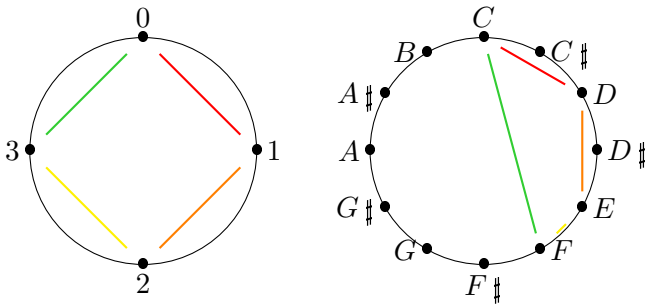
$$\sigma_{\alpha,\beta} = \rho_{\alpha,\beta} \circ \sigma_{0,0}$$

which simply means that, for any musical bar $\mathcal{B}$, $\sigma_{\alpha,\beta} \cdot \mathcal{B}$ is the reversed musical bar $-\mathcal{B}$ with an α-permutation of the onsets and a β-transposition of the pitches.

3) In this context, we use the words *rotations* and *reflections* because the maps we just defined simply correspond to the **dihedral group**, which acts on two polygons at the same time: the attack-times and the pitch-classes sets of the notes of the considered musical bar. In music theory, it corresponds to the generalized group "T/I" of musical translations and inversions, but we prefer to keep the dihedral vocabulary and notation to avoid confusion with quotient groups. In fact, we can still think of a bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ as a pair of (not necessarily convex) polygons on the circles of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, and it seems quite natural to use the action of the dihedral group on these two polygons. For example, recall the musical bar $\mathcal{B}$ of the above example, we had

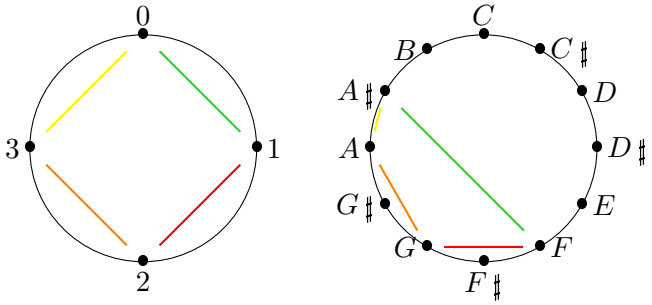
$$\mathcal{B} = \{(0,0),(1,2),(2,4),(3,5)\} \subset \mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z},$$

and $\mathcal{B}$ is also the following pair of polygons

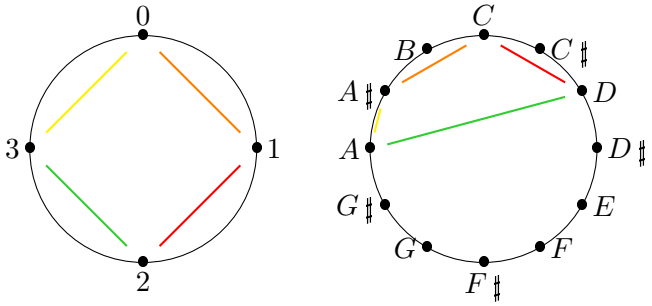


where we use color to underline the relationship between onsets and pitches. Then, we can represent the action of $\rho_{1,5}$ or $\sigma_{2,2}$ on $\mathcal{B}$ by the following corresponding drawings:

$$\rho_{1,5} \cdot \mathcal{B} = \{(0,10),(1,5),(2,7),(3,9)\}$$



$$\sigma_{2,2} \cdot \mathcal{B} = \{(0,10),(1,0),(2,2),(3,9)\}$$



We will now focus on the relationship between these well-defined rotations and reflections and the DFT-distance we defined in the first chapter (Definition 1.2.8). More precisely, we want to show that our metric is invariant under the action of the dihedral group or, in other words, that rotations and symmetries are isometries for our metric. We begin with the following lemma:

Lemma 2.4.4. *Let $\mathcal{B}$ be a musical bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. The magnitude of the DFT of $\mathcal{B}$ is invariant under rotations and reflections. More precisely, for any $(\alpha,\beta) \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we have*

i) $\mathcal{F}_{\rho_{\alpha,\beta}\cdot\mathcal{B}}(x,y) = \exp\left(\frac{-2i\pi\alpha x}{t}\right)\exp\left(\frac{-2i\pi\beta y}{p}\right) \times \mathcal{F}_{\mathcal{B}}(x,y)$

ii) $\mathcal{F}_{\sigma_{\alpha,\beta}\cdot\mathcal{B}}(x,y) = \exp\left(\frac{2i\pi\alpha x}{t}\right)\exp\left(\frac{2i\pi\beta y}{p}\right) \times \overline{\mathcal{F}_{\mathcal{B}}(x,y)}$

which directly implies that $\left|\mathcal{F}_{\rho_{\alpha,\beta}\cdot\mathcal{B}}\right| = |F_{\mathcal{B}}| = \left|\mathcal{F}_{\sigma_{\alpha,\beta}\cdot\mathcal{B}}\right|$.

Proof. Let $\rho_{\alpha,\beta}$ be a rotation of $\mathcal{B}$:

$$\begin{aligned}
\mathcal{F}_{\rho_{\alpha,\beta}\cdot\mathcal{B}}(x,y) &= \sum_{(k,l)\in\rho_{\alpha,\beta}\cdot\mathcal{B}} \exp\left(\tfrac{-2i\pi kx}{t}\right)\exp\left(\tfrac{-2i\pi ly}{p}\right)\\
&= \sum_{(k-\alpha,l-\beta)\in\mathcal{B}} \exp\left(\tfrac{-2i\pi kx}{t}\right)\exp\left(\tfrac{-2i\pi ly}{p}\right)\\
&= \sum_{(u,v)\in\mathcal{B}} \exp\left(\tfrac{-2i\pi(u+\alpha)x}{t}\right)\exp\left(\tfrac{-2i\pi(v+\beta)y}{p}\right)\\
&= \exp\left(\tfrac{-2i\pi\alpha x}{t}\right)\exp\left(\tfrac{-2i\pi\beta y}{p}\right)\sum_{(u,v)\in\mathcal{B}} \exp\left(\tfrac{-2i\pi ux}{t}\right)\exp\left(\tfrac{-2i\pi vy}{p}\right)\\
&= \exp\left(\tfrac{-2i\pi\alpha x}{t}\right)\exp\left(\tfrac{-2i\pi\beta y}{p}\right)\mathcal{F}_{\mathcal{B}}(x,y)
\end{aligned}$$

Now let $\sigma_{\alpha,\beta}$ be a reflection of $\mathcal{B}$, and we have similar calculations:

$$\begin{aligned}
\mathcal{F}_{\sigma_{\alpha,\beta}\cdot\mathcal{B}}(x,y) &= \sum_{(k,l)\in\sigma_{\alpha,\beta}\cdot\mathcal{B}} \exp\left(\tfrac{-2i\pi kx}{t}\right)\exp\left(\tfrac{-2i\pi ly}{p}\right)\\
&= \sum_{-(k+\alpha,l+\beta)\in\mathcal{B}} \exp\left(\tfrac{-2i\pi kx}{t}\right)\exp\left(\tfrac{-2i\pi ly}{p}\right)\\
&= \sum_{(u,v)\in\mathcal{B}} \exp\left(\tfrac{-2i\pi(-\alpha-u)x}{t}\right)\exp\left(\tfrac{-2i\pi(-\beta-v)y}{p}\right)\\
&= \exp\left(\tfrac{2i\pi\alpha x}{t}\right)\exp\left(\tfrac{2i\pi\beta y}{p}\right)\sum_{(u,v)\in\mathcal{B}} \exp\left(\tfrac{2i\pi ux}{t}\right)\exp\left(\tfrac{2i\pi vy}{p}\right)\\
&= \exp\left(\tfrac{2i\pi\alpha x}{t}\right)\exp\left(\tfrac{2i\pi\beta y}{p}\right)\overline{\mathcal{F}_{\mathcal{B}}(x,y)}
\end{aligned}$$

■

We can now prove the isometry theorem:

Theorem 2.4.5. *Let $\mathcal{B}$ and $\mathcal{B}'$ be two musical bars in $\mathbb{Z}/t\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}$. For any rotation ρ and reflection σ in $\mathbb{Z}/t\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}$, we have*

$$d_{\mathrm{DFT}}(\mathcal{B},\mathcal{B}') = d_{\mathrm{DFT}}\left(\rho\cdot\mathcal{B},\rho\cdot\mathcal{B}'\right) = d_{\mathrm{DFT}}\left(\sigma\cdot\mathcal{B},\sigma\cdot\mathcal{B}'\right).$$

which means that rotations and reflections are isometries on the metric set $\mathfrak{B}_{t,p}$ of all the musical bars together with the DFT-distance. In other words, the DFT-distance is invariant under the action of the dihedral group.

Proof. It all comes from previous Lemma 2.4.4 and Definition 1.2.8 of the metric $\mathrm{d}_{\mathrm{DFT}}$: for $\rho_{\alpha,\beta}$ a rotation in $\mathbb{Z}/t\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}$, we have

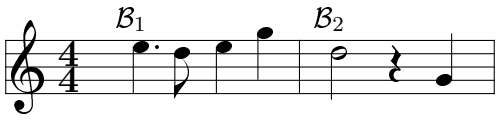
$$\begin{aligned}
\mathrm{d}_{\mathrm{DFT}}\left(\rho_{\alpha,\beta}\cdot\mathcal{B},\rho_{\alpha,\beta}\cdot\mathcal{B}'\right) &= d_1(M_{\rho_{\alpha,\beta}\cdot\mathcal{B}},M_{\rho_{\alpha,\beta}\cdot\mathcal{B}'})\\
&= \sum_{k=1}^{t}\sum_{l=1}^{p}\left|\mathcal{F}_{\rho_{\alpha,\beta}\cdot\mathcal{B}}(k,l)-\mathcal{F}_{\rho_{\alpha,\beta}\cdot\mathcal{B}'}(k,l)\right|\\
&= \sum_{k=1}^{t}\sum_{l=1}^{p}\left|\exp\left(\tfrac{-2i\pi\alpha k}{t}\right)\exp\left(\tfrac{-2i\pi\beta l}{p}\right)\right|\left|\mathcal{F}_{\mathcal{B}}(k,l)-\mathcal{F}_{\mathcal{B}'}(k,l)\right|\\
&= \sum_{k=1}^{t}\sum_{l=1}^{p}\left|\mathcal{F}_{\mathcal{B}}(k,l)-\mathcal{F}_{\mathcal{B}'}(k,l)\right|\\
&= \mathrm{d}_{\mathrm{DFT}}(\mathcal{B},\mathcal{B}')
\end{aligned}$$

Now for a reflection $\sigma_{\alpha,\beta}$:

$$\begin{aligned}
\mathrm{d_{DFT}}\left(\sigma_{\alpha,\beta}\cdot\mathcal{B},\sigma_{\alpha,\beta}\cdot\mathcal{B}'\right) &= \sum_{k=1}^{t}\sum_{l=1}^{p}\left|\mathcal{F}_{\sigma_{\alpha,\beta}\cdot\mathcal{B}}(k,l)-\mathcal{F}_{\sigma_{\alpha,\beta}\cdot\mathcal{B}'}(k,l)\right| \\
&= \sum_{k=1}^{t}\sum_{l=1}^{p}\left|\exp\left(\tfrac{2i\pi\alpha x}{t}\right)\exp\left(\tfrac{2i\pi\beta y}{p}\right)\left(\overline{\mathcal{F}_{\mathcal{B}}(k,l)}-\overline{\mathcal{F}_{\mathcal{B}'}(k,l)}\right)\right| \\
&= \sum_{k=1}^{t}\sum_{l=1}^{p}\left|\overline{\mathcal{F}_{\mathcal{B}}(k,l)}-\overline{\mathcal{F}_{\mathcal{B}'}(k,l)}\right| \\
&= \sum_{k=1}^{t}\sum_{l=1}^{p}\left|\overline{\mathcal{F}_{\mathcal{B}}(k,l)-\mathcal{F}_{\mathcal{B}'}(k,l)}\right| \\
&= \mathrm{d_{DFT}}(\mathcal{B},\mathcal{B}')
\end{aligned}$$

■

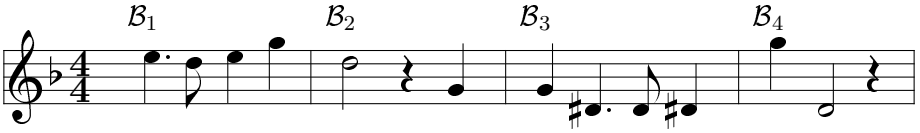
Examples 2.4.6. Let us consider the following score $\mathcal{S}=\{\mathcal{B}_1,\mathcal{B}_2\}$



which has the following description in $\mathbb{Z}/8\mathbb{Z}\times\mathbb{Z}/24\mathbb{Z}$

$$\begin{aligned}
\mathcal{B}_1 &= \{(0,15),\ (3,14),\ (4,15),\ (6,19)\} \\
\mathcal{B}_2 &= \{(0,14),\ (6,7)\}
\end{aligned}$$

1) We take the rotation $\rho_{2,12}$ that shifts $\mathcal{B}_1$ and $\mathcal{B}_2$ with a quarter note and translates the pitch-classes set by an octave. We then complete $\mathcal{S}$ with $\mathcal{B}_3=\rho_{2,12}\cdot\mathcal{B}_1$ and $\mathcal{B}_4=\rho_{2,12}\cdot\mathcal{B}_2$ and get the following score:



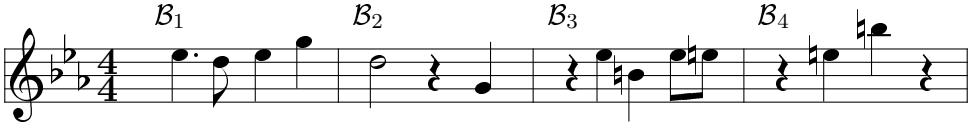
with $\mathcal{B}_3$ and $\mathcal{B}_4$ described by

$$\begin{aligned}
\mathcal{B}_3 &= \{(0,7),(2,3),(5,2),(6,3)\} \\
\mathcal{B}_4 &= \{(0,19),(2,2)\}
\end{aligned}$$

This resulting new score satisfies the relation

$$\mathrm{d_{DFT}}(\mathcal{B}_1,\mathcal{B}_2)=\mathrm{d_{DFT}}(\mathcal{B}_3,\mathcal{B}_4)\simeq 414.648$$

2) Now let us take the reflection $\sigma_{2,6}$ which inerts $\mathcal{B}_1$ and $\mathcal{B}_2$ and then adds a quarter note on attack-times sets and transposes the pitch-classes one by half on octave. We then get the following score:



with $\mathcal{B}_3 = \sigma_{2,6} \cdot \mathcal{B}_1$ and $\mathcal{B}_4 = \sigma_{2,6} \cdot \mathcal{B}_2$ described by

$$\begin{aligned} \mathcal{B}_3 &= \{(2,15),(4,11),(6,15),(7,16)\} \\ \mathcal{B}_4 &= \{(2,16),(4,23)\} \end{aligned}$$

This resulting new score satisfies the relation

$$\mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_1,\mathcal{B}_2) = \mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_3,\mathcal{B}_4) \simeq 414.648$$

We now give some musical applications of Theorem 2.4.5.

Corollary 2.4.7. *Let $\mathcal{B}$ be a musical bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. For every pair (α,β) in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we have*

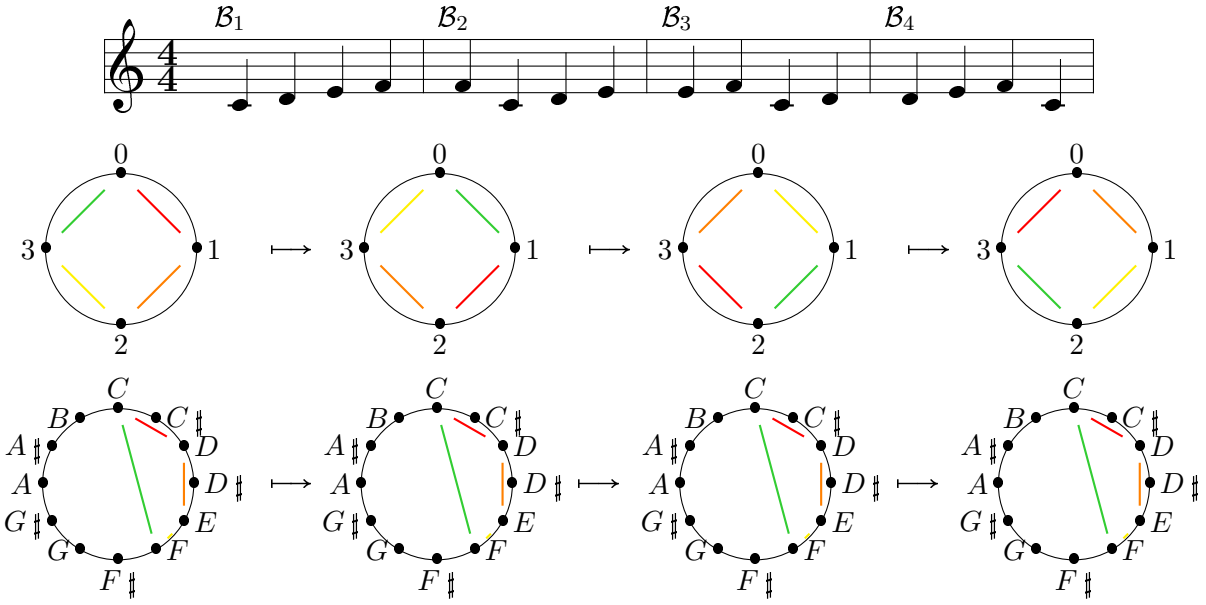
$$d_{\mathrm{DFT}}\left(\rho_{\alpha-1,\beta-1} \cdot \mathcal{B}, \rho_{\alpha,\beta} \cdot \mathcal{B}\right) = d_{\mathrm{DFT}}\left(\rho_{\alpha,\beta} \cdot \mathcal{B}, \rho_{\alpha+1,\beta+1} \cdot \mathcal{B}\right)$$

Proof. We just have to notice that $\rho_{1,1} \circ \rho_{\alpha,\beta} = \rho_{\alpha+1,\beta+1}$ and, since $\rho_{\alpha,\beta}$ is an isometry, it follows that

$$\mathrm{d}_{\mathrm{DFT}}\left(\rho_{\alpha-1,\beta-1} \cdot \mathcal{B}, \rho_{\alpha,\beta} \cdot \mathcal{B}\right) = \mathrm{d}_{\mathrm{DFT}}\left(\rho_{1,1} \circ \rho_{\alpha-1,\beta-1} \cdot \mathcal{B}, \rho_{1,1} \circ \rho_{\alpha,\beta} \cdot \mathcal{B}\right) = \mathrm{d}_{\mathrm{DFT}}\left(\rho_{\alpha,\beta} \cdot \mathcal{B}, \rho_{\alpha+1,\beta+1} \cdot \mathcal{B}\right).$$

■

Examples 2.4.8. This corollary has a particular musical meaning, especially when we exactly apply t times the same rotation $\rho_{1,0}$ to a musical bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, that means all the cyclic permutations until we get back to the initial musical bar. For instance, the following score is constructed in this way

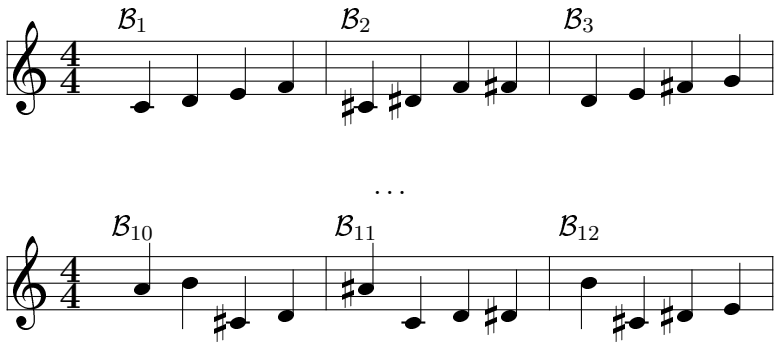


and we have the successive relations

$$\mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_1,\mathcal{B}_2) = \mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_2,\mathcal{B}_3) = \mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_3,\mathcal{B}_4) = \mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_1,\mathcal{B}_4) \simeq 98,886.$$

Obviously, we also have $\mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_1,\mathcal{B}_3) = \mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_2,\mathcal{B}_4) \simeq 82,359$ because the same rotation $(\rho_{1,0})$ is applied: $\rho_{1,0} \cdot \mathcal{B}_1 = \mathcal{B}_2$ and $\rho_{1,0} \cdot \mathcal{B}_3 = \mathcal{B}_4$.

On the same idea, we could choose to apply the same transposition p times on the set of pitch-classes, that is, apply the rotation $\rho_{0,1}$ on $\mathcal{B}_1$ until get back to the initial bar:



In this case, $\rho_{0,1} \cdot \mathcal{B}_n = \mathcal{B}_{n+1}$ for any $n \in \{1, \dots, 11\}$ and so we get that

$$d_{\mathrm{DFT}}(\mathcal{B}_1, \mathcal{B}_2) = d_{\mathrm{DFT}}(\mathcal{B}_2, \mathcal{B}_3) = \dots = d_{\mathrm{DFT}}(\mathcal{B}_{10}, \mathcal{B}_{11}) = d_{\mathrm{DFT}}(\mathcal{B}_{11}, \mathcal{B}_{12}) = d_{\mathrm{DFT}}(\mathcal{B}_{12}, \mathcal{B}_1) \simeq 110,978$$

2.4.2. Conjugate musical bar

Recall Lemma 2.4.4: let $\mathcal{B} \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ be a musical bar and take the particular case where $\alpha = \beta = 0$. We thus have

$$\mathcal{F}_{\rho_{0,0} \cdot \mathcal{B}}(x, y) = \mathcal{F}_{\mathcal{B}}(x, y) \text{ and } \mathcal{F}_{\sigma_{0,0} \cdot \mathcal{B}}(x, y) = \overline{\mathcal{F}_{\mathcal{B}}(x, y)}$$

and it turns out that taking a $(0,0)$-rotation does not change the Fourier coefficients, but a $(0,0)$-reflection is the same as taking the conjugated associated matrix. Therefore, we define the **conjugate bar** of $\mathcal{B}$, denoted by $\overline{\mathcal{B}}$, the musical bar which is given by

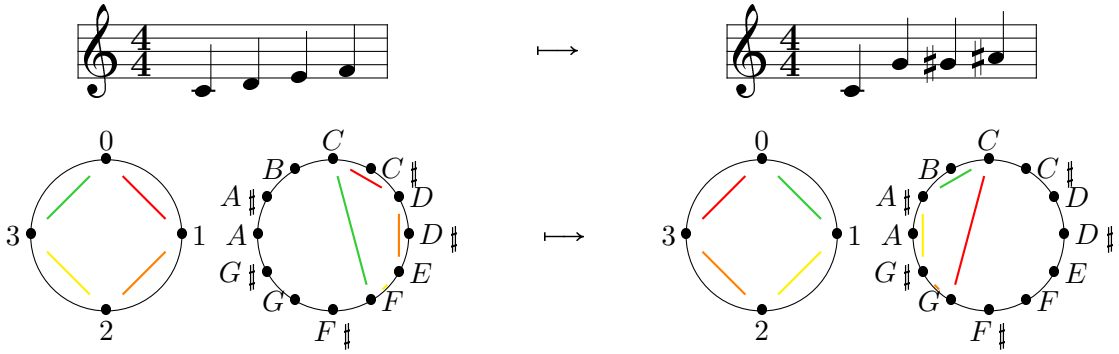
$$\overline{\mathcal{B}} = \sigma_{0,0} \cdot \mathcal{B}\,.$$

The Fourier coefficients for $\overline{\mathcal{B}}$ are the conjugate of those for $\mathcal{B}$:

$$\mathcal{F}_{\overline{\mathcal{B}}}(x, y) = \overline{\mathcal{F}_{\mathcal{B}}(x, y)} \text{ for all } (x, y) \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}\,.$$

Example 2.4.9. We take back the musical bar $\mathcal{B} = \{(0,0), (1,2), (2,4), (3,5)\}$ from Example 2.4.2. The corresponding conjugate musical bar of $\mathcal{B}$ is given by

$$\overline{B} = \sigma_{0,0} \cdot \mathcal{B} = \{(0,0), (3,10), (2,8), (1,7)\} = \{(0,0), (1,7), (2,8), (3,10)\}$$



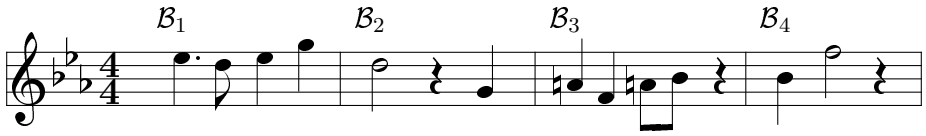
The following theorem directly comes from the isometry Theorem 2.4.5 for reflections:

Theorem 2.4.10. *The DFT-distance is invariant under the complex conjugate: for two musical bars $\mathcal{B}$ and $\mathcal{B}'$, we have*

$$d_{\mathrm{DFT}}(\mathcal{B}, \mathcal{B}') = d_{\mathrm{DFT}}(\overline{\mathcal{B}}, \overline{\mathcal{B}'})$$

In other words, conjugate a musical bar is an isometry for the DFT-distance.

Example 2.4.11. Recall the score $\mathcal{S} = \{\mathcal{B}_1, \mathcal{B}_2\}$ from Example 2.4.6, with time-unit 8 and pitch-unit 24. We decide to complete this score by computing $\overline{\mathcal{B}_1} = \sigma_{0,0} \cdot \mathcal{B}_1$ and $\overline{\mathcal{B}_2} = \sigma_{0,0} \cdot \mathcal{B}_2$, which gives us:



with $\overline{\mathcal{B}_1}$ and $\overline{\mathcal{B}_2}$ described by

$$\begin{aligned} \mathcal{B}_3 &= \{(0,9),(2,5),(4,9),(5,10)\} \\ \mathcal{B}_4 &= \{(0,10),(2,17)\} \end{aligned}$$

This new score thus satisfies the relation

$$\mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_1, \mathcal{B}_2) = \mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_3, \mathcal{B}_4) \simeq 414.648$$

2.4.3. Homotheties

We will now define an other type of musical transformation.

Definition 2.4.12. Let $\mathcal{B}$ be a musical bar in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ and (α, β) be a pair of **invertible** elements in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. The **homothety with ratio** (α, β), or **(α, β)-homothety** of $\mathcal{B}$ is a map $\lambda_{\alpha,\beta} : \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ such that, for each pair (x, y) in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we have

$$\lambda_{\alpha,\beta}(x, y) = (\alpha x, \beta y)$$

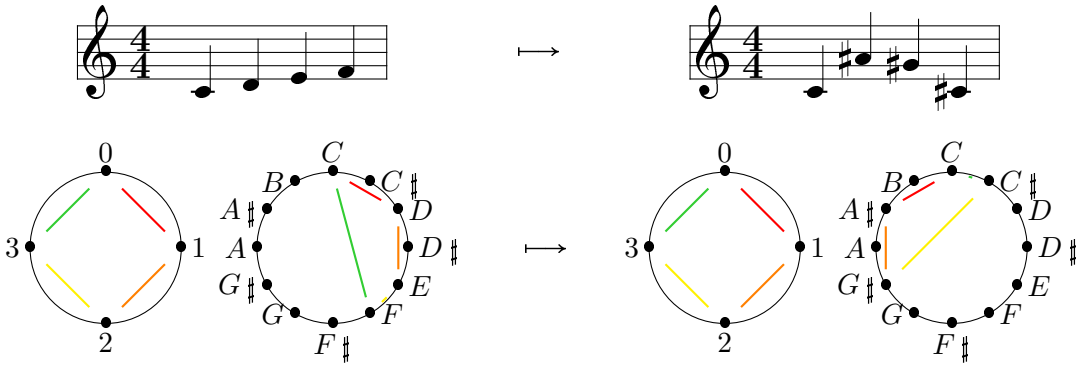
and thus we define

$$\lambda_{\alpha,\beta} \cdot \mathcal{B} = \{(\alpha t, \beta p),\ (t, p) \in \mathcal{B}\}.$$

Examples 2.4.13.

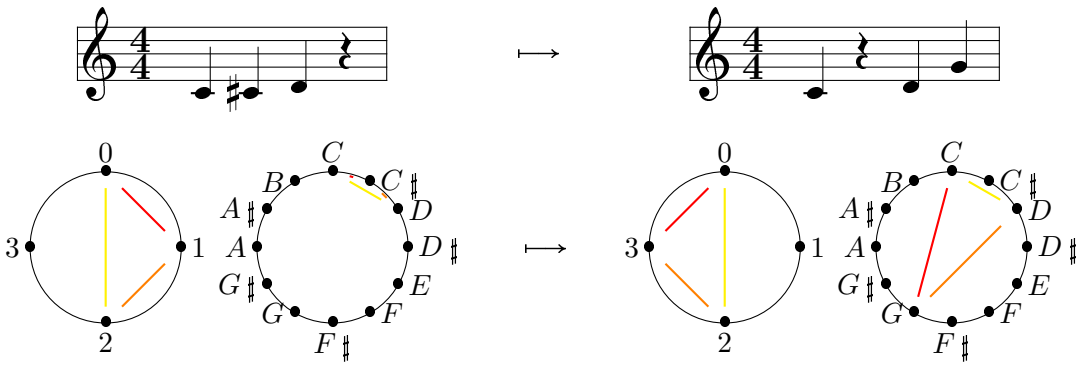
1) Let $\mathcal{B} = \{(0,0),(1,2),(2,4),(3,5)\}$ be the musical bar from Example 2.4.2, with time-unit 4 and pitch-unit 12. For $\alpha = 1$ and $\beta = 5$, $\lambda_{1,5}$ is a $(1,5)$-homothety of $\mathcal{B}$ that preserves the onsets and extends the pitches by a fourth, and we have

$$\lambda_{1,5} \cdot \mathcal{B} = \{(0,0),(1,10),(2,8),(3,1)\} :$$



2) Let $\mathcal{B} = \{(0,0),(1,1),(2,2)\}$, still in $\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$. For $\alpha = 3$ and $\beta = 7$, $\lambda_{3,7}$ is a $(3,7)$-homothety of $\mathcal{B}$ and we have

$$\lambda_{3,7} \cdot \mathcal{B} = \{(0,0),(3,7),(2,2)\} = \{(0,0),(2,2),(3,7)\} :$$



We also have an isometry theorem in the case of homothety:

Theorem 2.4.14. *Let $\mathcal{B}$ and $\mathcal{B}'$ be two musical bars in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. If $\lambda_{\alpha,\beta}$ is a (α,β)-homothety in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ with (α,β)* ***invertible*** *elements, we have*

$$d_{\mathrm{DFT}}(\mathcal{B},\mathcal{B}') = d_{\mathrm{DFT}}\left(\lambda_{\alpha,\beta}\cdot\mathcal{B}, \lambda_{\alpha,\beta}\cdot\mathcal{B}'\right)$$

which means that homotheties are isometries on the set $\mathfrak{B}_{t,p}$ of all the musical bars with the DFT-distance.

Proof. Let $(x,y) \in \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. We can write the Fourier coefficients of $\lambda_{\alpha,\beta}\cdot\mathcal{B}$ in term of $\mathcal{F}_{\mathcal{B}}$, and because (α,β) is invertible in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we have the following description:

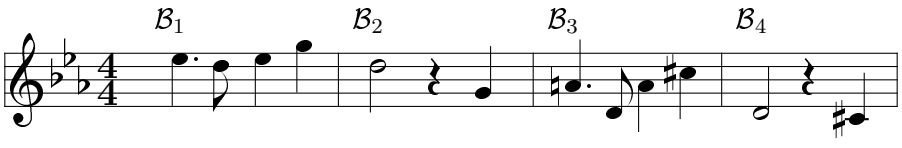
$$\begin{aligned}
\mathcal{F}_{\lambda_{\alpha,\beta}\cdot\mathcal{B}}(x,y) &= \sum_{(k,l)\in\lambda\cdot\mathcal{B}} \exp\left(\tfrac{-2ik\pi x}{t}\right)\exp\left(\tfrac{-2il\pi y}{p}\right)\\
&= \sum_{(k',l')\in\mathcal{B}} \exp\left(\tfrac{-2i\alpha k'\pi x}{t}\right)\exp\left(\tfrac{-2i\beta l'\pi y}{p}\right)\\
&= \mathcal{F}_{\mathcal{B}}(\alpha x, \beta y)
\end{aligned}$$

Now because (α,β) is invertible in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, the distance is the same so we finally get that

$$\begin{aligned}
\mathrm{d}_{\mathrm{DFT}}(\lambda_{\alpha,\beta}\cdot\mathcal{B}, \lambda_{\alpha,\beta}\cdot\mathcal{B}') &= \sum_{k=1}^{t}\sum_{l=1}^{p}\left|\mathcal{F}_{\lambda_{\alpha,\beta}\cdot\mathcal{B}}(k,l) - \mathcal{F}_{\lambda_{\alpha,\beta}\cdot\mathcal{B}'}(k,l)\right|\\
&= \sum_{k'=1}^{t}\sum_{l'=1}^{t}\left|\mathcal{F}_{\mathcal{B}}(\alpha k',\beta l') - \mathcal{F}_{\mathcal{B}'}(\alpha k',\beta l')\right|\\
&= \mathrm{d}_{\mathrm{DFT}}(\mathcal{B},\mathcal{B}')
\end{aligned}$$

■

Example 2.4.15. Recall the score $\mathcal{S} = \{\mathcal{B}_1, \mathcal{B}_2\}$ from Example 2.4.6, with time-unit 8 and pitch-unit 24. If we apply the homothety of a quaver note and fifth $h_{1,7}$ on $\mathcal{B}_1$ and $\mathcal{B}_2$, we get



with $\mathcal{B}_3 = \lambda_{1,7}\cdot\mathcal{B}_1$ and $\mathcal{B}_4 = \lambda_{1,7}\cdot\mathcal{B}_2$ described by

$$\begin{aligned}
\mathcal{B}_3 &= \{(0,9),(3,2),(4,9),(6,13)\}\\
\mathcal{B}_4 &= \{(0,2),(6,1)\}
\end{aligned}$$

The resulting new scores thus satisfies the relation

$$\mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_1,\mathcal{B}_2) = \mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_3,\mathcal{B}_4) \simeq 414.648.$$

PART II

Persistent homology on musical bars

Chapter 3.

Mathematical background

In this chapter, we will introduce persistent homology, which is the mathematical tool we are going to use to apply the process of Topological Data Analysis to musical analysis. We will begin with a brief reminder of simplicial homology, and then move on to the definitions of filtration, persistence and barcodes. Notice that all computations will be done in $\mathbb{F}_2$, the ground field with two elements.

3.1. Simplicial homology

In this first section, we simply give an overview of simplicial theory and the basic elements we will need in order to define persistent homology. We refer to [28] for more details on this topic.

Definition 3.1.1. A **simplicial complex** is a pair (K, V) where V is a set of **vertices** and $K \subset \mathcal{P}(V)$ is a set of **simplices**, such as

i) $V = \bigcup_{\sigma \in K} \sigma$

ii) if $\sigma \subset K$ and $\tau \subset \sigma$, then $\tau \in K$.

Most of the time, we will refer to K as a simplicial complex by omitting the set of vertices.

If $\sigma \in K$ is a simplex of K and $\tau \subset \sigma$, we say that τ is a **face** of σ. The dimension of $\sigma \in K$ is given by its number of vertices: $\dim \sigma = |\sigma| - 1$. If $\dim \sigma = n$, we then say that σ is a n**-simplex**. The dimension of a simplicial complex K is given by $\dim K = \sup_{\sigma \in K} \dim \sigma$.

We can always think of a simplex by representing its **geometric realization**: a n-simplex is realized by the convex hull of $n+1$ affinely independent points in $\mathbb{R}^d$, for any $d \geq n$. Then, the geometric realization of a simplicial complex is given by the realization of each simplex where the simplices are glued together on common faces (see Figure 3.1.1).

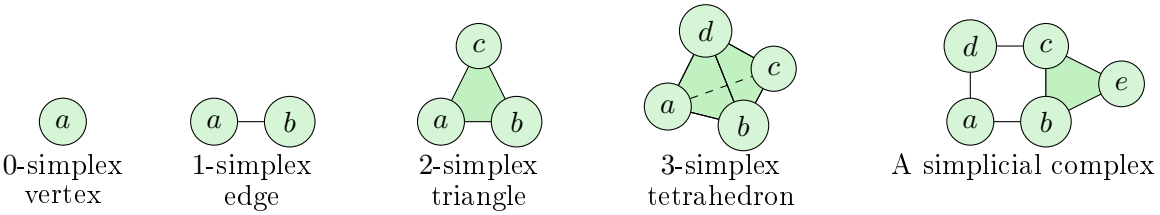


Figure 3.1.1: The geometric realization of low-dimensional n-simplices for $n \in \{0, 1, 2, 3\}$ (left) and a simplicial complex (V, K) of dimension 2: the set of vertices is $V = \{a, b, c, d, e\}$ and the largest simplex is the triangle $\{b, c, e\}$.

Definition 3.1.2. Let (V_1, K_1) and (V_2, K_2) be two simplicial complexes. A **simplicial morphism** between these two complexes is an application defined on the set of vertices

$$f : V_1 \to V_2$$

such as, if $\sigma \in K_1$, then $f(\sigma) \in K_2$ (with $f(\sigma) = \{f(v) \mid v \in \sigma\}$).

Furthermore, if $f : (V_1, K_1) \to (V_2, K_2)$ is a simplicial morphism, there exists a linear application $\mathbb{R}[V_1] \to \mathbb{R}[V_2]$ induced by f and which sends $|(V_1, K_1)|$ on $|(V_2, K_2)|$. We denote by $|f|$ the resulting continuous map:

$$|f| : |(V_1, K_1)| \to |(V_2, K_2)| .$$

Now we want to build an algebraic structure on any simplicial complex, so that we can define the simplicial homology of a complex.

Definitions 3.1.3. Let K be a simplicial complex.

1. The *nth*-**chain group** $C_n(K, \mathbb{F}_2) = C_n(K)$ of K is the $\mathbb{F}_2$-vector space generated by the n-simplices of K. An element of $C_n(K)$ is called a n-**chain** $c \in C_n(K)$ and is given by the *formal sum*
$$c = \sum_\sigma \lambda_\sigma \sigma, \text{ with } \lambda_\sigma \in \mathbb{F}_2 .$$

2. The *nth*-**boundary operator** $\partial_n : C_n(K) \to C_{n-1}(K)$ is a morphism linearly defined on a chain $c = \sum_\sigma \lambda_\sigma \sigma \in C_n(K)$ by its action on its n-simplices:
$$\partial_n(\sigma) = \sum_i \{v_0, v_1, \dots, \hat{v_i}, \dots, v_n\} \quad \text{and} \quad \partial_n(c) = \lambda_\sigma \sum_\sigma \partial_n(\sigma)$$

The most important thing with the boundary operator is the following fundamental lemma:

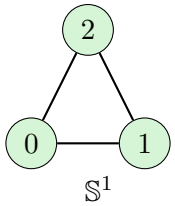
Lemma 3.1.4. *For any n, we have $\partial_n \circ \partial_{n+1} = 0$.*

This property means that, for any simplicial complex K, there exists a sequence of chain groups with boundary operators

$$\dots \xrightarrow{\partial_{n+2}} C_{n+1}(K) \xrightarrow{\partial_{n+1}} C_n(K) \xrightarrow{\partial_n} C_{n-1}(K) \xrightarrow{\partial_{n-1}} \dots$$

where for each dimension n, $\operatorname{im} \partial_{n+1} \subset \ker \partial_n$. In other words, for any simplicial complex K, we have a family of chain groups $\big(C(K)_n\big)_{n \in \mathbb{N}}$ together with a family of boundary operators $(\partial_n)_{n \in \mathbb{N}}$ that satisfies the above inclusion. This kind of structure is called a **chain complex**.

Example 3.1.5. As an illustration of these notions, we can describe all the chain groups for the one-dimensional sphere $\mathbb{S}^1$:



$\mathbb{S}^1$

Indeed, $\dim \mathbb{S}^1 = 1$ so $C_n(\mathbb{S}^1) = \{0\}$ for any $n < 0$ and $n \geq 2$. Therefore, we just need to look at the set of vertices and the set of edges to get that

$$\begin{array}{rcccc} C_0(\mathbb{S}^1) & = & \mathbb{F}_2 \cdot \langle 0 \rangle \oplus \mathbb{F}_2 \cdot \langle 1 \rangle \oplus \mathbb{F}_2 \cdot \langle 2 \rangle & \simeq & {\mathbb{F}_2}^3 \\ C_1(\mathbb{S}^1) & = & \mathbb{F}_2 \cdot \langle 01 \rangle \oplus \mathbb{F}_2 \cdot \langle 12 \rangle \oplus \mathbb{F}_2 \cdot \langle 02 \rangle & \simeq & {\mathbb{F}_2}^3 \end{array}$$

As an example of a 1-chain, we can consider the outline of $\mathbb{S}^1$, which is given by the sum

$$c = \langle 01\rangle + \langle 12\rangle - \langle 02\rangle.$$

Moreover, the only non zero boundary operator is $\partial_1 : C_1(\mathbb{S}^1) \to C_0(\mathbb{S}^1)$, and its action on any edge of $\mathbb{S}^1$ is simply given by

$$\partial_1\langle 01\rangle = \langle 1\rangle + \langle 0\rangle, \quad \partial_1\langle 12\rangle = \langle 2\rangle + \langle 1\rangle, \quad \partial_1\langle 02\rangle = \langle 2\rangle + \langle 0\rangle.$$

Finally, we can look at ∂_1 applied on the above 1-chain $c = \langle 01\rangle + \langle 12\rangle - \langle 02\rangle$, and we have

$$\partial_1\big(\langle 01\rangle + \langle 12\rangle + \langle 02\rangle\big) = \big(\langle 1\rangle + \langle 1\rangle\big) + \big(\langle 0\rangle + \langle 0\rangle\big) + \big(\langle 2\rangle + \langle 2\rangle\big) = 0.$$

Notations: Let K be a simplicial complex, and $(\partial_n)_{n\in\mathbb{N}}$ be its associated family of boundary operators. For each $n \in \mathbb{N}$, we denote by $Z_n(K)$ and $B_n(K)$ respectively the kernel and the image of ∂_n and ∂_{n+1}, i.e.

$$Z_n(K) = \ker \partial_n \text{ and } B_n(K) = \operatorname{im} \partial_{n+1}.$$

We call $Z_n(K)$ and $B_n(K)$ the n**-cycles** and the n**-boundaries** of K, respectively. Therefore, by fundamental Lemma 3.1.4, we have $B_n(K) \subset Z_n(K)$. This inclusion leads to the definition of simplicial homology.

Definition 3.1.6. The *nth*-**homology group** of K is the quotient

$$H_n(K, \mathbb{F}_2) = H_n(K) := {}^{Z_n(K)}\big/_{B_n(K)}.$$

This is a $\mathbb{F}_2$-vector space and its dimension is called the *nth*-**Betti number**:

$$\beta_n(K) = \dim H_n(K).$$

Example 3.1.7. For instance, let us compute the homology of $\mathbb{S}^1$. Recall from Example 3.1.5 the following sequence of chain groups

$$0 \xrightarrow{\partial_2} C_1(\mathbb{S}^1) \xrightarrow{\partial_1} C_0(\mathbb{S}^1) \xrightarrow{\partial_0} 0$$

First, we have $B_1(\mathbb{S}^1) = \operatorname{im} \partial_2 = 0$ so $H_1(\mathbb{S}^1) = {}^{Z_1(\mathbb{S}^1)}\big/_{B_1(\mathbb{S}^1)} = \ker \partial_1$. Moreover, any element of $\ker \partial_1$ is generated by the cycle $\langle 01\rangle + \langle 12\rangle + \langle 02\rangle$, and so

$$H_1(\mathbb{S}^1) \simeq \mathbb{F}_2 \text{ and } \beta_1(\mathbb{S}^1) = 1.$$

On the other hand, $H_0(\mathbb{S}^1) = {}^{Z_0(\mathbb{S}^1)}\big/_{B_0(\mathbb{S}^1)}$ and since $\dim\ker \partial_1 = 1$, $\dim B_0(\mathbb{S}^1) = \dim\operatorname{im} \partial_1 = 2$, so $Z_0(\mathbb{S}^1) = \ker \partial_0 = C_0(\mathbb{S}^1)$. We thus get that

$$H_0(\mathbb{S}^1) \simeq \mathbb{F}_2 \text{ and } \beta_0(\mathbb{S}^1) = 1.$$

Finally, we obviously have $H_n(\mathbb{S}^1) = 0$ in any other case. We can also generalize this result for any dimensional sphere and any ground field: indeed, for any n, any m and any field k we have

$$H_n(\mathbb{S}^m, k) = \begin{cases} k \text{ if } n = 0 \text{ or } m \\ 0 \text{ otherwise} \end{cases}$$

3.2. Persistent homology

The idea now is to generalize these previous notions by computing simplicial homology for an entire filtration of simplicial complex. The purpose is to study the topological features that *persist* along time of filtration. Again, we give here an overview of persistent homology and we refer to [59] for more details on this topic.

Definition 3.2.1. A **filtered simplicial complex** $K = (K^i)_i$ is a nested sequence of simplicial complexes, such as

$$\varnothing = K^{-1} \subset K^0 \subseteq K^1 \subseteq \ldots \subseteq K^N \subseteq \ldots$$

We say that the filtration is **finite** if the sequence is eventually constant:

$$\varnothing = K^{-1} \subset K^0 \subseteq K^1 \subseteq \ldots \subseteq K^N = K$$

In that case, we say that $\{0, 1, \ldots, N\}$ is the set of **filtration times**.

In all this manuscript, we will always consider **finite** filtered simplicial complexes, with $N+1$ times of filtration. As an illustration of this definition, Figure 3.2.1 represents a filtered complex with six times of filtration.

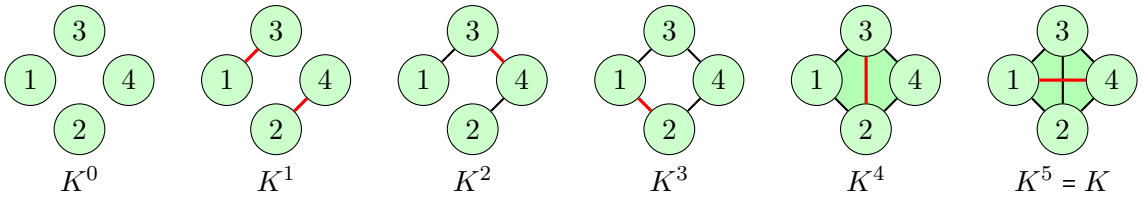


Figure 3.2.1: A filtered simplicial complex K with six times of filtration.

Let $(i, p) \in \mathbb{N}^2$ be a couple of integers. For any filtered complex K, we have a natural inclusion map

$$\eta^{i,p} : K^i \hookrightarrow K^{i+p}$$

that induces a homological morphism

$$\eta_*^{i,p} : H_*(K^i) \to H_*(K^{i+p}).$$

Each one of the maps $\eta_*^{i,p}$ sends a homological class to the one containing it. Therefore, for each degree n, we get the following sequence of nth-homology groups:

$$0 \longrightarrow H_n(K^0) \xrightarrow{\eta_n^{0,1}} H_n(K^1) \xrightarrow{\eta_n^{1,1}} \ldots H_n(K^{N-1}) \xrightarrow{\eta_n^{N-1,1}} H_n(K^N) = H_n(K)$$

Thus, the image of each map $\eta_*^{i,p}$ will describe the homology evolution during the filtration, which is exactly the goal of persistent homology.

Definitions 3.2.2. Let K be a filtered complex, and i, p, n be integers in $\mathbb{N}$.

1. The **p-persistent homology group** associated with K^i in degree n is given by

$$H_n^{i,p}(K, \mathbb{F}_2) = H_n^{i,p}(K) := \operatorname{im} \eta_n^{i,p} : K^i \to K^{i+p}.$$

2. The **p-persistent Betti number** associated with K^i is the dimension of $H_n^{i,p}$:

$$\beta_n^{i,p}(K) := \dim H_n^{i,p}(K).$$

Remark 3.2.3. For each pair of integers $(n,i) \in \mathbb{N}^2$, there is a boundary operator $\partial_n^i : C_n(K^i) \to C_{n-1}(K^i)$ so we can also define cycles, boundaries and homology groups as follows:

$$Z_n^i(K) = Z_n(K^i),\ B_n^i(K) = B_n(K^i) \text{ and } H_n^i(K) = H_n(K^i).$$

Thus, an equivalent and perhaps more intuitive definition of persistent homology is given by the following quotient:

$$H_n^{i,p}(K) = {}^{Z_n^i}\!/_{Z_n^i \cap B_n^{i+p}}.$$

In fact, by the first definition, an element of $H_n^{i,p}(K)$ is a generate at time i of the filtration that is still not a boundary after p times of filtration. This is exactly the meaning of the quotient above. In other words, the purpose is to look for homology classes that persist along the filtration, that means cycles that do not become boundaries.

From [59], we have the following structure theorem of persistent homology:

Theorem 3.2.4. *Let K be a filtered simplicial complex. In any degree n, we have the following isomorphism of $\mathbb{F}_2[t]$-module:*

$$\bigoplus_i H_n(K^i) \cong \Big(\bigoplus_i t^{a_i} \cdot \mathbb{F}_2[t]\Big) \oplus \Big(\bigoplus_j t^{b_j} \cdot {}^{\mathbb{F}_2[t]}\!/_{(t^{c_j})}\Big) \tag{3.2.1}$$

Theorem's equivalence 3.2.1 encodes the homology for a whole filtration of a simplicial complex K in a single formula. Moreover, it shows that there are two families of homology generators: indeed, we have classes that arise at time a_i of the filtration and never die (left side of 3.2.1, also called the *free part*), and we have the others that arise at time b_j and die at time $b_j + c_j$ (right side of 3.2.1, also called the *torsion part*). Notice that, in the case of finite filtration, there are no free part in 3.2.1. This observation naturally leads to the construction of **barcodes** (or persistence diagrams), which are graphics that provide a natural visualization of persistent homology.

Definition 3.2.5. Let K be a filtered simplicial complex and $d \in \mathbb{N}$ be a fixed degree. The **barcode** $\mathrm{BC}_d(K)$ associated with $H_d(K)$ is the graph where the x-axis describes the time of filtration and each generator of $H_d(K)$ corresponds to a *bar* whose start and end are given by Theorem's equivalence 3.2.1:

- a class that was born at time a_i and never dies is a bar that starts at the abscissa point a_i and never stops
- a class that was born at time b_j and died at time $b_j + c_j$ is a bar that starts and ends at abscissa points b_j and $b_j + c_j$, respectively.

Remark 3.2.6. The above equivalence is cited in general case, but in practice we will only consider **finite** filtration (see Definition 3.2.1): therefore, there is no free part and Theorem's equivalence 3.2.1 simply becomes

$$\bigoplus_i H_n(K^i) \cong \bigoplus_j t^{b_j} \cdot {}^{\mathbb{F}_2[t]}\!/_{(t^{c_j})} \tag{3.2.2}$$

Consequently a barcode $\mathrm{BC}_d(K)$ associated with a finite filtered complex and a fixed degree d contains only finite bars. For $d = 0$, there is always at least one bar of the form $[0, N+1]$, where $\{0, 1, \ldots, N\}$ is the set of filtration times.

An illustration of Theorem 4.2 and Definition of barcodes 3.2.5 is given in Figure 3.2.2, where we computed the associated family of barcodes with filtration from Figure 3.2.1: in degree 0 (left), we see that there are 4 bars at time 0, and the first two stop when the edges $\{13\}$ and

$\{24\}$ appear, while the third stops when the filtration becomes connected. After that, there is only one last bar which never stops and which corresponds to the only connected component of the filtered complex. This description corresponds to the following equivalence in degree 0 according to Remark 3.2.6:

$$\bigoplus_{i=0}^{5} H_0(K^i) \cong \left(\mathbb{F}_2[t]/_{t}\right) \oplus \left(\mathbb{F}_2[t]/_{t}\right) \oplus \left(\mathbb{F}_2[t]/_{t^2}\right) \oplus \left(\mathbb{F}_2[t]/_{t^5}\right)$$

In degree 1 (right), there is a single bar corresponding to the cycle that appears at time 3 of the filtration, when the outline of the square is completed. This cycle becomes a boundary at time 4, when the square is filled. The barcode in degree 1 corresponds to the following isomorphism:

$$\bigoplus_{i=0}^{6} H_1(K^i) \cong t^3 \cdot \mathbb{F}_2[t]/_{t}$$

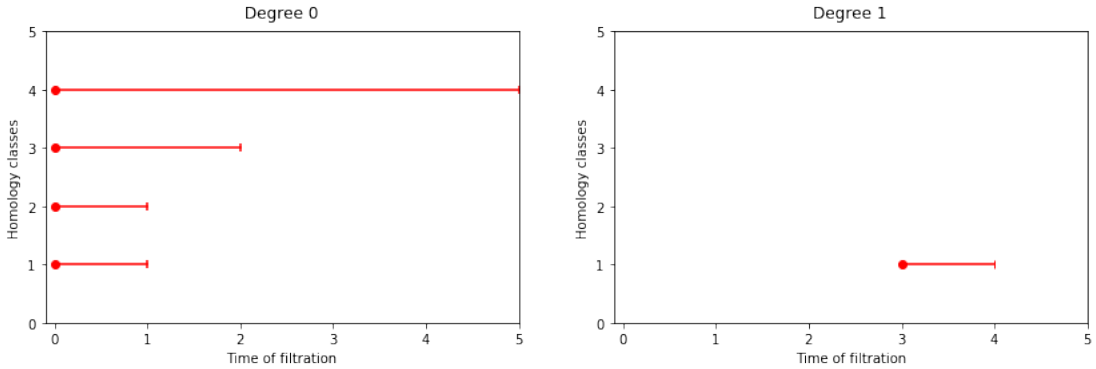


Figure 3.2.2: The associated barcodes with filtration from Figure 3.2.1.

Moreover, we can easily give a generator for each bar of this family of barcodes: in degree 1, there is only one bar and we just saw that it corresponds to the cycle

$$\{13\} + \{34\} + \{42\} + \{21\}.$$

In degree 0, we have the following sequence of homology groups

$$H_0(K^0) \longrightarrow H_0(K^1) \longrightarrow H_0(K^2) \longrightarrow H_0(K^3) \longrightarrow H_0(K^4) \longrightarrow H_0(K^5)$$

and each arrow sends a class in degree 0 to a class containing it. For instance, between time 0 and time 1, there are two generators that disappear so we look for two elements of $H_0(K^0)$ that become boundaries for K^1: since the edges $\{13\}$ and $\{24\}$ appear at time 1 of the filtration, the generators for the first two bars are given by

$$\{1\} + \{3\} \quad \text{and} \quad \{2\} + \{4\}.$$

In the same way, the third bar corresponds to the birth of the edge $\{34\}$, so a generator of this bar is given by the element

$$\{3\} + \{4\} \in C_0(K^2).$$

Finally, the last bar is given by a homology class that rises at time 0 and never dies, so it simply corresponds to any element of $C_0(K^0)$, that means for instance the vertex $\{1\}$.

Barcodes are commonly used in the context of Topological Data Analysis, where the general purpose is to associate a family of such graphs with a starting object, and then study and use them as a kind of topological signature. This is what we will set up in the next sections.

Chapter 4.

Musical scores and filtration

In this chapter we show in what context we use persistent homology for musical analysis. More specifically, we are going to explain how we turn a musical score into a filtration, in order to extract a topological signature from it using barcodes. This is what we call Topological Data Analysis, which we will introduce in the first section. We will then show how we apply this process to musical scores, using the DFT as a metric.

4.1. Topological Data Analysis

Before using persistent homology in the context of musical analysis, we must understand in what way this tool can be used to *analyze* any starting object (see [26] to have an overview of the possible applications): in practice, there is an object to study with a hidden topological structure that potentially allows us to *understand* it, in a sense that remains to be precise. For this purpose, the object is transformed into a **point cloud**, that is to say into a finite metric set. From this point cloud, we build a filtered simplicial complex using the Vietoris-Rips algorithm, which will be presented below. Once this is done, it is possible to compute persistent homology and then extract the associated family of barcodes. Therefore, we can use this family as a **topological fingerprint** to characterize the starting object, as we will see in applications Part III. This general TDA process is summarized in Figure 4.1.1.

Figure 4.1.1: Topological Data Analysis process.

The first step is not trivial, and obviously there are several arbitrary choices we can make, depending on the starting object and what we want to obtain. In this work, we will study musical pieces, and more precisely associated musical scores, represented by a corresponding MIDI file. From this type of object, we will extract a point cloud using the musical bars of the score, as this is presented in the next Section 4.2. The other step that needs to be clarified is the arrow that links the point cloud together with the filtered complex. Here we have chosen to use the **Vietoris-Rips method**, which is presented in the following definition.

Definition 4.1.1. Let $X = \{x_1, ..., x_N\}$ be a point cloud, that means a collection of points in a metric space, and let $\epsilon \geq 0$ be a non-negative parameter. The **Vietoris-Rips complex** $\mathcal{R}_\epsilon(X)$ is the simplicial complex where :

i) X is the set of vertices

ii) $\sigma = \{x_1, ..., x_n\}$ is a n-simplex if and only if the vertices it contains are pairwise close, i.e. if $d(x_i, x_j) \leq \epsilon$ for all pairs x_i, x_j of σ.

Remark 4.1.2. Notice that a complex obtained in this way has the distinctive property that it is determined by its 1-skeleton (i.e. the induced graph): indeed, a collection of vertices of the complex spans a simplex if and only if they form clique (that is, a complete subgraph) of the 1-skeleton. Therefore, there is no loss of information by focusing on the 1-skeleton instead of the whole complex.

Let us look at Figure 4.1.2 as an illustration of this definition. One can immediately notice that if two parameters ϵ and ϵ' are chosen such as $\epsilon < \epsilon'$, there is a simple inclusion $R_\epsilon(X) \hookrightarrow R_{\epsilon'}(X)$. With this observation, we understand how the filtration arises from this construction: let us take $X = \{x_1, ..., x_N\}$ as our starting point cloud and consider a very small value for ϵ. For instance, $R_0(X)$ is the 0-dimensional simplicial complex with N connected components. Also by simply increasing the parameter ϵ, we will get a natural sequence of nested simplicial complexes, that means a filtered simplicial complex. Also notice that when we talk about "time of filtration", we implicitly assume that we choose each parameter in a discrete set, and that the time is discretized. Here we refer to [27] for details on Vietoris-Rips filtration and Topological Data Analysis.

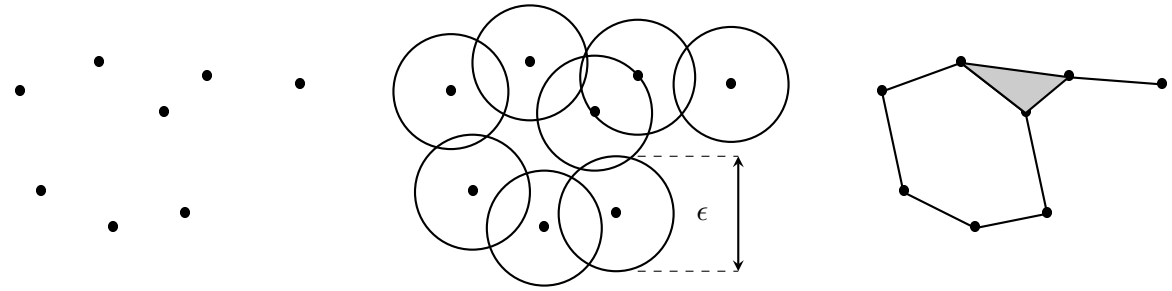


Figure 4.1.2: A fixed point cloud X (on the left) from which we extract a simplicial complex $R_\epsilon(X)$ (on the right) for a given paramater ϵ using the Vietoris-Rips method.

We understand here that the most important question now is the choice of the point cloud, and especially the distance we will use to compare all the points together. This question will be discussed in the next section.

Remark 4.1.3. In the musical applications, we will essentially use the filtered complexes produced by the Vietoris-Rips method, since this is the most common practice in Topological Data Analysis. In Chapter 6 and more precisely in Section 6.4, we will adapt this filtration construction by associating a "graph-type" to any musical score, using Remark 4.1.2.

4.2. From a score to a filtered complex

In this section, we will use what we just introduced, and also the first Part I of this manuscript. In fact, recalling TDA process from Figure 4.1.1, in the very first step we have to make a choice, that is, which point cloud we are going to extract from our starting musical piece $\mathfrak{P}$. Following the idea from Section 1.2, we will represent our musical piece we are working with by a score $\mathcal{S}_\mathfrak{P}$. For implementation, we will also turn $\mathcal{S}_\mathfrak{P}$ into a MIDI file.

Let $\mathfrak{P}$ be any musical piece and $\mathcal{S}_\mathfrak{P}$ be the score representing it. In Section 1.2 and more precisely Definition 1.2.2, we defined the notion of **musical bar** of a score: if t is the **time unit** and p the **ambitus**, then a musical bar $\mathcal{B}$ is of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ and an element of $\mathcal{B}$ is a note which

is characterized by its **attack-time** modulo t and its **pitch-class** modulo p. Thus, a **musical score** $\mathcal{S}$ becomes simply the set of its distinct musical bars modulo (t,p).

Now let $\mathcal{S}_{\mathfrak{P}} = \{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_N\}$ be the associated score with $\mathfrak{P}$, where $\mathcal{B}_i \neq \mathcal{B}_j$ for $i \neq j$ and N is the number of distinct musical bars of $\mathfrak{P}$. By Definition 1.2.5, each musical bar $\mathcal{B}_j \in \mathcal{S}_{\mathfrak{P}}$ is represented by a matrix of Fourier coefficients $M_{\mathcal{B}_j}$, where

$$M_{\mathcal{B}_j} = \Big(\mathcal{F}_{\mathcal{B}_j}(x,y)\Big)_{(x,y)\in\mathbb{Z}/t\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}} \quad \text{and} \quad \mathcal{F}_{\mathcal{B}_j}(x,y) = \sum_{(k,l)\in\mathcal{B}_j} \exp\big(\frac{-2i\pi kx}{t}\big)\exp\big(\frac{-2i\pi ly}{p}\big).$$

Starting from there, we want to create a metric on the set of all the musical bars, as we previously mentioned in the first part of this manuscript (recall Definition 1.2.8): indeed, all we have to do is to compare the associated matrix of Fourier coefficients together, that means simply choose a metric under the set of matrices of $\mathcal{M}_{t\times p}(\mathbb{C})$. For computational constraints, we choose to work with the metric that comes from the 1-norm, as in the following definition:

Definition 4.2.1. Let $A = (a_{ij})_{\substack{1\leq i\leq m\\ 1\leq j\leq n}}$ be an element of $\mathcal{M}_{m\times n}(\mathbb{C})$. The 1-norm of A is defined as follow:

$$\| A \|_1 = \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|.$$

If A and B are two elements of $\mathcal{M}_{m\times n}(\mathbb{C})$, then the **distance** between A and B is given by

$$\| A - B \|_1 .$$

The above definition provides a natural metric under the set of all the musical bars $\mathfrak{B}_{t,p}$:

Definition 4.2.2. Let $\mathcal{B}$ and $\mathcal{B}'$ be two elements of $\mathfrak{B}_{t,p}$. The **DFT-distance** between $\mathcal{B}$ and $\mathcal{B}'$ is given by the 1-distance between their respective Fourier coefficients matrices:

$$\mathrm{d}_{\mathrm{DFT}}(\mathcal{B},\mathcal{B}') = \| M_{\mathcal{B}} - M_{\mathcal{B}'} \|_1 = \sum_{x=1}^{t}\sum_{y=1}^{p} |\mathcal{F}_{\mathcal{B}}(x,y) - \mathcal{F}_{\mathcal{B}'}(x,y)|$$

We denoted by $(\mathfrak{B}_{t,p}, \mathrm{d}_{\mathrm{DFT}})$ the metric space of all the musical bars of $\mathbb{Z}/t\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}$ equipped with the DFT-distance.

We are now able to turn any musical piece $\mathfrak{P}$ into a point cloud, that means a subset of a metric set. The following definition summarizes the process, and we will use it in most of the applications of Part III.

Definition 4.2.3. Let $\mathfrak{P}$ be a musical piece and $\mathcal{S}_{\mathfrak{P}} = \{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_N\}$ the score that represents it. Then, $\mathcal{S}_{\mathfrak{P}}$ is a subset of the metric space $(\mathfrak{B}_{t,p}, \mathrm{d}_{\mathrm{DFT}})$, and the **associated point cloud** with $\mathfrak{P}$ is built in the following way:

i) Each **point** is a musical bar $\mathcal{B}_i$.

ii) The **distance** between two musical bars $\mathcal{B}_i$ and $\mathcal{B}_j$ is given by the DFT-distance, that means $\mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_i,\mathcal{B}_j) = \| M_{\mathcal{B}_i} - M_{\mathcal{B}_j} \|_1$ where $M_{\mathcal{B}_i}$ and $M_{\mathcal{B}_j}$ are the respective matrix of Fourier coefficients.

Example 4.2.4. To illustrate the previous definitions, let us consider the score $\mathcal{S}$ from Figure 4.2.1 which is excerpt from the music piece *One summer's day*, composed by Joe Hisaishi in 2001 for the movie *Spirited Away*.

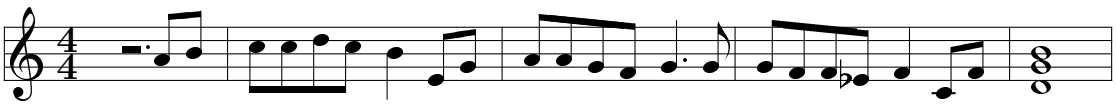


Figure 4.2.1: The score $\mathcal{S}$ extracted from *One summer's day* written by Joe Hisaishi in 2001 for the movie *Chihiro* of the same year.

The above score is divided into five distinct musical bars: $\mathcal{S} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \mathcal{B}_4, \mathcal{B}_5\}$. Furthermore, the shortest note here is a quaver one, so the time unit will be $t = 8$. Also the lowest note used is a C_4, the C on the fourth octave of the piano, while the highest one is a D_5. We then cross two octaves in this score, so the pitch unit will be $p = 24$, with $C_4 = 0 = C_6$, $C_5 = 12$ and $B_5 = 23$. With these observations, we have for each $\mathcal{B}_i \in \mathcal{S}$:

$$\mathcal{B}_i \in \mathbb{Z}/8\mathbb{Z} \times \mathbb{Z}/24\mathbb{Z}$$

By Definition 1.2.2, we are now able to give a description for each musical bar $\mathcal{B}_i$:

$$\begin{aligned}
\mathcal{B}_1 &= \{(6,9),\ (7,11)\} \\
\mathcal{B}_2 &= \{(0,12),\ (1,12),\ (2,14),\ (3,12),\ (4,11),\ (6,4),\ (7,7)\} \\
\mathcal{B}_3 &= \{(0,9),\ (1,9),\ (2,7),\ (3,5),\ (4,7),\ (7,7)\} \\
\mathcal{B}_4 &= \{(0,7),\ (1,5),\ (2,5),\ (3,3),\ (4,5),\ (6,0),\ (7,5)\} \\
\mathcal{B}_5 &= \{(0,2),\ (0,7),\ (0,11)\}
\end{aligned}$$

For each one of the $\mathcal{B}_i$, there is a matrix M_i of Fourier coefficients, which provides a natural way to compare the musical bars of $\mathcal{S}$ together, using Definition 4.2.2. Figure 4.2.1 gives the table of these computing distances.

	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_3$	$\mathcal{B}_4$	$\mathcal{B}_5$
$\mathcal{B}_1$	0	515.114	488.053	492.526	387.27
$\mathcal{B}_2$		0	550.136	605.697	550.149
$\mathcal{B}_3$			0	572.54	523.797
$\mathcal{B}_4$				0	476.545
$\mathcal{B}_5$					0

Table 4.2.1: The table of distances between musical bars of $\mathcal{S}$ from *One summer's day*.

At this moment, we have exactly our expected point cloud, where the points are $\{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \mathcal{B}_4, \mathcal{B}_5\}$ and the metric is given by the distances from Table 4.2.1. We can then construct the associated filtered complex using the Vietoris-Rips method from Definition 4.1.1. To do this, we discretize time using Table 4.2.1: each distance is a time of filtration that corresponds to a parameter ϵ and, for each value, we build the associated complex $R_\epsilon(\mathcal{S})$, by adding a new edge for the corresponding pair of bars. In this example, the different values of ϵ are given by

$$\{0,\ 387.27,\ 476.545,\ 488.053,\ 492.526,\ 515.114$$

$$523.797,\ 550.136,\ 550.149,\ 572.541,\ 605.697\}$$

For instance, for a parameter $\epsilon = 387.27$, $R_\epsilon(\mathcal{S})$ is the complex with $\mathcal{S} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \mathcal{B}_4, \mathcal{B}_5\}$ as the set of vertices and only one edge $(\mathcal{B}_1, \mathcal{B}_5)$. Then, for $\epsilon = 476.545$, we simply add the edge $(\mathcal{B}_4, \mathcal{B}_5)$ to the previous complex, and so on. We thus obtain a filtration, which is shown in Figure 4.2.2, where each complex is represented by its corresponding graph (only vertices and edges) for ease of illustration.

Once we have the filtration, we are able to compute persistent homology, that is, the associated family of barcodes. Here and throughout this paper, we will focus only on barcodes in degree 0 and degree 1: indeed, the intuition of a musical interpretation goes with these low dimensions, as for instance we think of a cycle in dimension 1 as a musical loop. Of course, higher dimensions could be interpreted in future works. Figure 4.2.3 shows associated family of barcodes with *One summer's day* in degree 0 and degree 1. We can notice that there is no element of homology in degree 1, which is not that surprising considering the number of musical bars in the score.

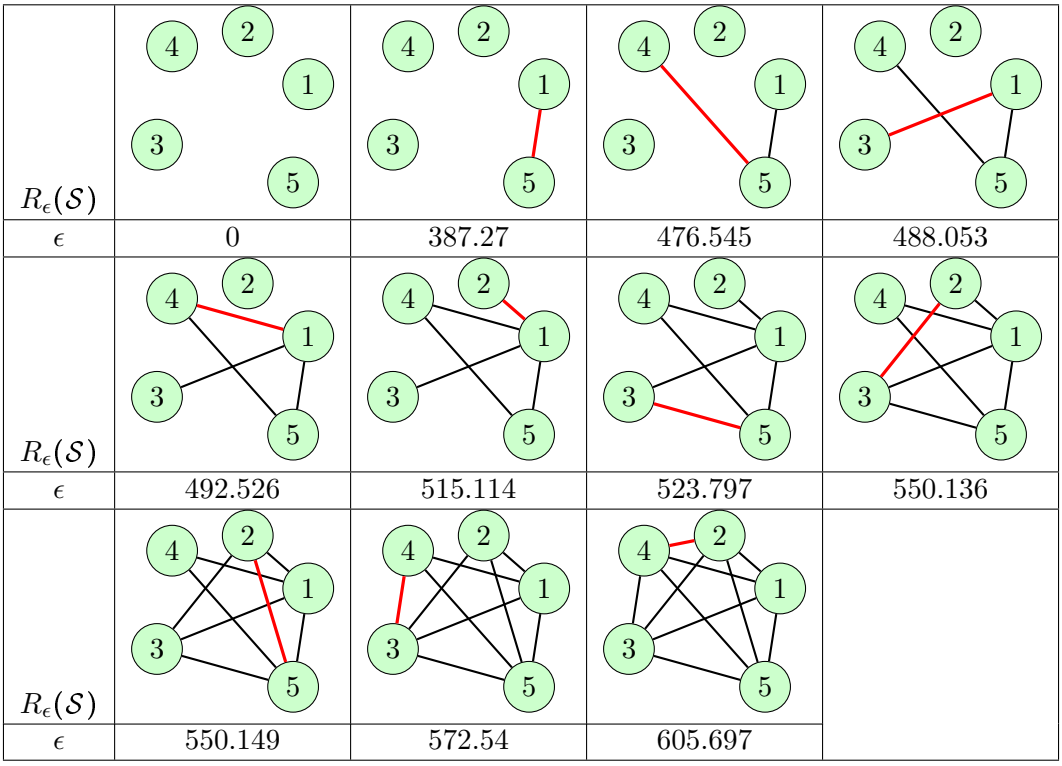


Figure 4.2.2: The filtered simplicial complex associated with score $\mathcal{S}$ from *One summer's day*: each distance from Table 4.2.1 is used as a discretization of time and we represent each complex of the filtration by its associated graph.

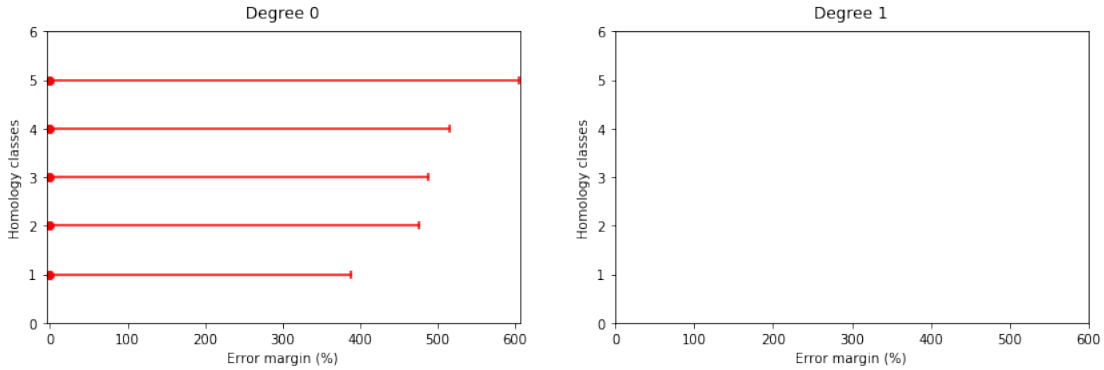


Figure 4.2.3: The family of barcodes (in degrees 0 and 1) associated with the score $\mathcal{S}$ from *One summer's day*.

4.3. Normalization and scaling parameter

In Example 4.2.4 of the previous section, the time was discretized in order to compute the Vietoris-Rips complex associated with $\mathcal{S}$, and the different values of the parameter ϵ were taken into a set of finite number of distances. In particular, for two different scores we might have several different values of ϵ, and the resulting family of barcodes might have completely different scales. However, one of the main purposes of this thesis is to compare different scores with each other, that means to compare different filtered complexes and families of barcodes. Therefore, it would be reasonable to ask that all the barcodes to be written in the same scale.

Following on this observation, we make a choice here that we will keep for the rest of this paper: for any given musical score $\mathcal{S}_{\mathfrak{P}}$ associated with any given musical piece $\mathfrak{P}$, the resulting filtration will be encoded with discretized time values of the form $\epsilon = t\rho$, where $t \in \{0, 1, \ldots, 100\}$ and ρ is a fixed constant. At this point, each filtration will have the same number of complexes and each associated barcodes will be onthe same scale, more precisely between 0 and 100. Actually, the fixed constant ρ is computed as follows: for a given score $\mathcal{S} = \{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$, let us set $d_{\max}$ as the maximum distance of all the values of ϵ. We then define ρ as the **precision** we want to work with, that is:

$$\rho = \frac{d_{\max}}{100}$$

Thus for each $t \in \{0, 1, \ldots, 100\}$ we consider the associated Vietoris-Rips complex $\mathcal{R}_{t\rho}(\mathcal{S})$. More precisely, for each value of $\epsilon_{i,j} = \mathrm{d_{DFT}}(\mathcal{B}_i, \mathcal{B}_j)$, which corresponds to a distance between two musical bars $\mathcal{B}_i, \mathcal{B}_j$, we associate a value $t_{i,j} \in \{0, 1, \ldots, 100\}$ given by

$$t_{i,j} = \begin{cases} 1 & \text{if } \epsilon_{i,j} < 1 \\ \left\lfloor \frac{\epsilon_{i,j}}{\rho} \right\rfloor & \text{if } \epsilon_{i,j} \geq 1 \end{cases}$$

Furthermore, instead of talking about "time t of the filtration", we will now say that we are looking at the complex on the musical piece on with a **scaling parameter** (or sometimes an **error-margin**) of t%. Indeed, we think of the presence of an edge between two musical bars as an indication that they are "similar", and the parameter t controls the way in which we choose to make this rigorous: for a small value of t, there are few edges which means that the bars are finely distinguished, while for t large enough we allow coarser identifications.

As an application of this remark, in the previous Example 4.2.4 we have $d_{\max} = 605.697$ so our precision-constant ρ is given by

$$\rho = 6.05697$$

If we take $\epsilon_{1,3} = 488.053$, we get that the corresponding $t \in \{0, 1, \ldots, 100\}$ is given by

$$t_{1,3} = \left\lfloor \frac{\epsilon_{1,3}}{\rho} \right\rfloor = \left\lfloor \frac{488.053}{6.05697} \right\rfloor = 80$$

so the edge $(\mathcal{B}_1, \mathcal{B}_3)$ comes at 80% of the filtration. Figure 4.3.1 gives the new table of distances after normalization. The last edge $(\mathcal{B}_2, \mathcal{B}_4)$ comes at the end of the filtration, that means with a scale of 100%. Furthermore, Figure 4.3.1 gives the corresponding filtration (illustrated with graphs only), where one can see that, with a scale of 90%, there are two edges that appears: $(\mathcal{B}_2, \mathcal{B}_5)$ and $(\mathcal{B}_2, \mathcal{B}_3)$. This is due to the fact that the distances computed with the DFT-distance were really closed, so our normalization simply combines them. Finally, Figure 4.3.2 gives the new associated family of barcodes, with the scale between 0% and 100%.

	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_3$	$\mathcal{B}_4$	$\mathcal{B}_5$
$\mathcal{B}_1$	0%	85%	80%	81%	63%
$\mathcal{B}_2$		0%	90%	100%	90%
$\mathcal{B}_3$			0%	94%	86%
$\mathcal{B}_4$				0%	78%
$\mathcal{B}_5$					0%

Table 4.3.1: The table of distances between musical bars of $\mathcal{S}$ from *One summer's day* after normalization.

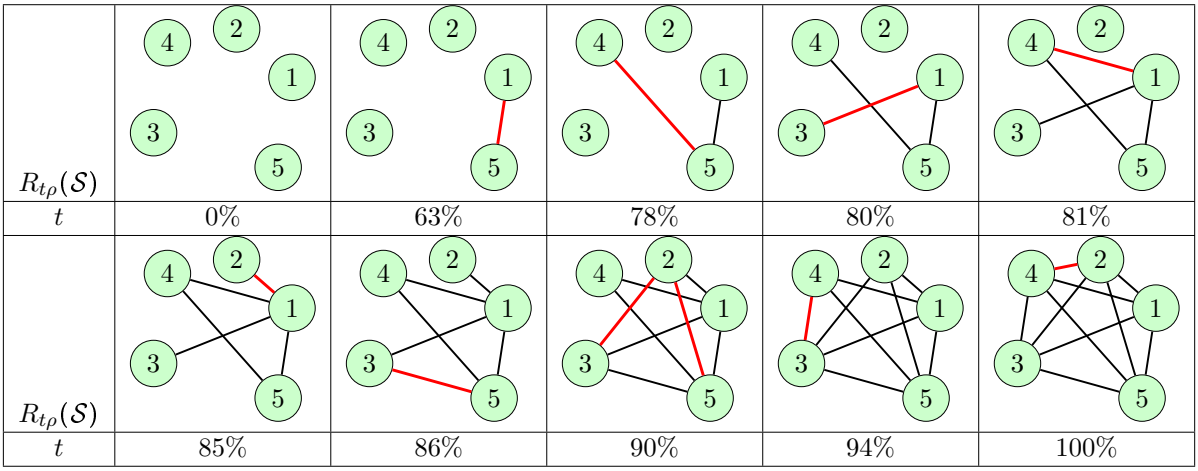


Figure 4.3.1: The filtered simplicial complex associated with score $\mathcal{S}$ from *One summer's day* after normalization.

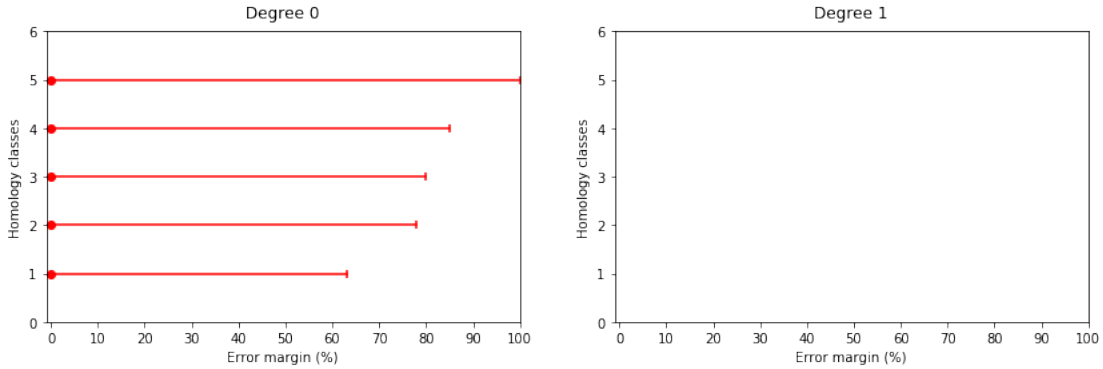


Figure 4.3.2: The family of barcodes associated with the score $\mathcal{S}$ from *One summer's day* after normalization. The barcodes scales are between 0% and 100%.

PART III

Musical applications

Chapter 5.

The DFT as a metric on the set of notes and chords

In this chapter, our aim is to understand in depth how the DFT works as a *metric*, and in particular to answer the question of whether this new metric makes musical sense. To do this, we will try it on some artificially created musical scores: more precisely, recall that in our Definition 1.2.2 a score is given by a non-ordered set of musical bars. In fact, we can construct a musical score made of very simple musical bars, such as bars containing only a N-chord. In this case, the pitches are all given with onset 0, so we take $t = 1$ as a unit of time. In particular, this means that we are back in the one-dimensional case. However, we can use the DFT to measure the distance between two given N-chords, and to do this we compute barcodes that lead us to understand the connection between the musical bars.

For this analysis, we are going to build three types of musical scores:

- the set of all the musical bars made up of a single arbitrary note from $\mathbb{Z}/12\mathbb{Z}$ (the chromatic scale)
- the set of all the musical bars made up of a single note taken from a given scale (diatonic, pentatonic)
- the set of all the musical bars made up of a 3-chords from a given family, such as the major and minor chords (the Euler's Tonnetz)

In the first case, the results are inconclusive but the method provides an order between the intervals of the chromatic scale based on the DFT. For specific sets of musical bars such as scales or Tonnetze, the filtrations are much more interesting: for a given scaling parameter, we recover some shapes that are characteristic of the studied scores (the torus for the Euler's Tonnetz, for instance). We also see the PLR-group, which arises naturally from this construction. With these results, the metric based on the DFT together with persistent homology seem to be some worthwhile tools to highlight the structure of a musical piece.

5.1. A metric on the set of notes

The model described in Chapter 4 is now reduced to a simpler case: a **note** can be seen as a musical bar of the form

$$\mathcal{B}_n = \{(0, n)\} \subset \mathbb{Z}/1\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$$

where n is the pitch-class associated with the note in $\mathbb{Z}/p\mathbb{Z}$. In this chapter, we will be working most of the time in one octave, that means for $p = 12$. Also notice that, for this model, we are somehow back in the one-dimensional case: in fact, using Definition 1.2.5, each musical bar $\mathcal{B}_n$ corresponds to a family of Fourier coefficients $\left(\mathcal{F}_{\mathcal{B}_n}(y)\right)_{y\in\{0,1,\dots,11\}}$ of the simple form:

$$\mathcal{F}_{\mathcal{B}_n}(y) = \exp\left(\frac{-2i\pi n}{12}\right)^y \tag{5.1.1}$$

In other words, each musical bar of one single note with pitch-class $n \in \mathbb{Z}/12\mathbb{Z}$ has a corresponding matrix of Fourier coefficients $M_n \in \mathcal{M}_{1\times 12}(\mathbb{C})$ of the above form. This means that by using the DFT-distance of Definition 4.2.2, we are able to compute the distance between any two notes of respective pitches n_a and n_b, and this is reduced to the following simple calculation:

$$\mathrm{d}_{\mathrm{DFT}}(n_a, n_b) = \mathrm{d}_{\mathrm{DFT}}(\mathcal{B}_{n_a}, \mathcal{B}_{n_b}) = \sum_{y=1}^{12} \left|\exp\left(\frac{-2i\pi n_a y}{12}\right) - \exp\left(\frac{-2i\pi n_b y}{12}\right)\right| \tag{5.1.2}$$

In this case, we can apply our model to any musical scale by converting it into a score and then into a point cloud (see Definition 4.2.3): therefore, we can compute and analyze the corresponding family of barcodes. In the following examples, we will compare different scales to understand the musical meaning of Distance 5.1.2, and what the different filtrations and barcodes bring to the study.

Furthermore, with the model given by the DFT and Equation 5.1.1, each musical pitch in $\mathbb{Z}/12\mathbb{Z}$ is associated with a family of Fourier coefficients that are given by the successive powers of a $12th$ root of unity. We set up this correspondence in Table 5.1.1, which can be useful for calculations.

Label	Pitch (in $\mathbb{Z}/12\mathbb{Z}$)	Fourier coefficients	Label	Pitch (in $\mathbb{Z}/12\mathbb{Z}$)	Fourier coefficients
C	0	$(1^y)_{y=0,1,\ldots,11}$	$F\sharp$	6	$\left((-1)^y\right)_{y=0,1,\ldots,11}$
$C\sharp$	1	$\left(e^{\frac{-i\pi y}{6}}\right)_{y=0,1,\ldots,11}$	G	7	$\left(e^{\frac{5i\pi y}{6}}\right)_{y=0,1,\ldots,11}$
D	2	$\left(e^{\frac{-i\pi y}{3}}\right)_{y=0,1,\ldots,11}$	$G\sharp$	8	$\left(e^{\frac{2i\pi y}{3}}\right)_{y=0,1,\ldots,11}$
$D\sharp$	3	$\left((-i)^y\right)_{y=0,1,\ldots,11}$	A	9	$(i^y)_{y=0,1,\ldots,11}$
E	4	$\left(e^{\frac{-2i\pi y}{3}}\right)_{y=0,1,\ldots,11}$	$A\sharp$	10	$\left(e^{\frac{i\pi y}{3}}\right)_{y=0,1,\ldots,11}$
F	5	$\left(e^{\frac{-5i\pi y}{6}}\right)_{y=0,1,\ldots,11}$	B	11	$\left(e^{\frac{i\pi y}{6}}\right)_{y=0,1,\ldots,11}$

Table 5.1.1: Correspondence between pitch-classes in $\mathbb{Z}/12\mathbb{Z}$ and their family of Fourier coefficients.

5.1.1. The chromatic scale

Let us start with the chromatic scale, whose score is presented in Figure 5.1.1. This scale contains all the pitch-classes in $\mathbb{Z}/12\mathbb{Z}$, so here we are studying a score $\mathcal{S} = \{\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_{11}\}$ where for each n:

$$\mathcal{B}_n = \{(0, n)\} \subset \mathbb{Z}/1\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}.$$

Notice that we are still using the convention that C is 0 and B is 11, but everything can be done for any musical transposition. In score from Figure 5.1.1, we have chosen to represent each note by a whole note, but we could also use quarter notes or any other duration (only the onset matters here).

Using the distance between the notes from Table 5.1.1 and persistent homology, we can now compute the family of barcodes associated with the chromatic scale, which is shown in Figure 5.1.2. The advantage of such examples with a small number of musical bars is that we can also analyze the filtration itself, describe the cycles and compare the different distances. Therefore, the table of distances between all the notes of the chromatic scale is given in Table 5.1.2.

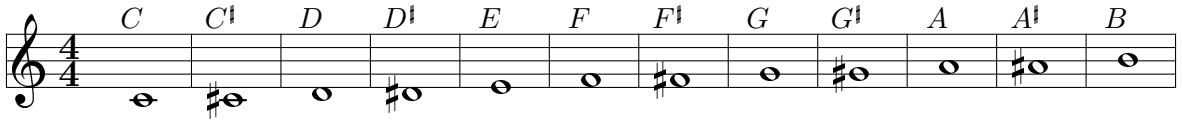


Figure 5.1.1: The chromatic scale with whole notes in $\mathbb{Z}/12\mathbb{Z}$.

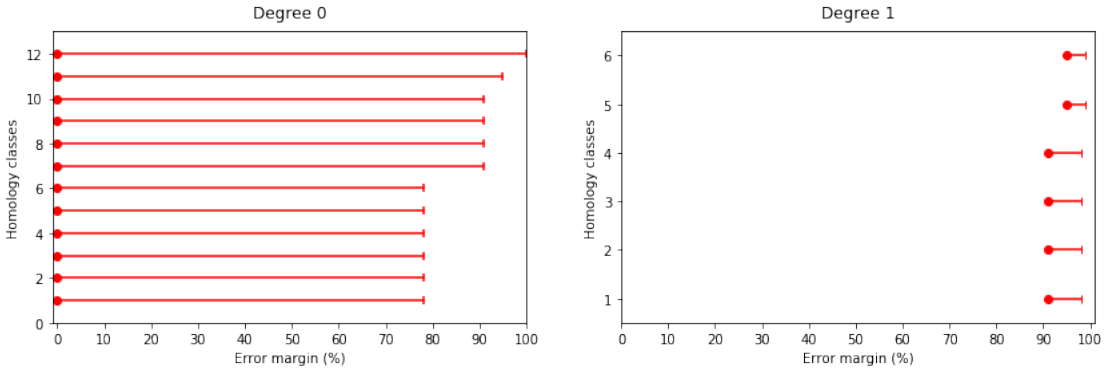


Figure 5.1.2: The associated family of barcodes with the chromatic scale.

	$F^\sharp$	E	$G^\sharp$	$D^\sharp$	A	D	$A^\sharp$	$C^\sharp$	F	G	B
C	12.0	13.856	13.856	14.485	14.485	14.928	14.928	15.192	15.192	15.192	15.192
0	+6	+4	−4	+3	−3	+2	−2	+1	+5	+7	+11

Table 5.1.2: The table of distances between the different notes of the chromatic scale. Here we have made the calculations starting with $C = 0$, but we can transpose them to any other pitch-class. The first row gives the distances computed with the DFT, and the second one gives the corresponding harmonic intervals.

Table 5.1.2 shows that the closest note of one given pitch-class (here C) is given by its augmented fourth (here $F^\sharp$). In fact, using Distance 5.1.2 and Table 5.1.1, we get that

$$\mathrm{d}_{\mathrm{DFT}}(C, F^\sharp) = \sum_{y=1}^{12} |1 - (-1)^y| = 12.$$

Furthermore, we can generalize this calculation for any pair of pitch-classes $(n_a, n_b) \in (\mathbb{Z}/12\mathbb{Z})^2$ using 5.1.2, and we get that 12 is actually the minimal DFT-distance between n_a and n_b:

$$\begin{aligned}
\mathrm{d}_{\mathrm{DFT}}(n_a, n_b) &= \sum_{y=1}^{12} \left| \exp\left(\tfrac{-2i\pi n_a y}{12}\right) - \exp\left(\tfrac{-2i\pi n_b y}{12}\right) \right| \\
&= \sum_{y=1}^{12} \left| 1 - \exp\left(\tfrac{-2i\pi (n_b - n_a) y}{12}\right) \right| \\
&\geq \left| \sum_{y=1}^{12} 1 - \exp\left(\tfrac{-2i\pi (n_b - n_a)}{12}\right)^y \right| \\
&= 12
\end{aligned}$$

because we have $\sum_{y=0}^{11} \omega^y = 0$ for any $12th$ root of unity ω. In other words, the closest pitch to a given one is always its augmented fourth. In terms of filtration and considering barcode in degree 0, at a scale of 78%, which is the first connection moment, there are exactly six connected

components rising at the same time. These components are given by the joining of the augmented fourths together, as it is shown in Figure 5.1.3.

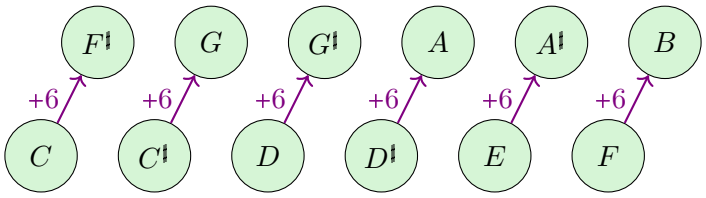


Figure 5.1.3: The filtration associated with the chromatic scale at a scale of 78%.

If we still focus on barcode in degree 0, we can see that the next step where the components of the filtration are gathering is given with a scaling parameter of 91%, and we now have two connected components. In fact, this corresponds to the moment when major thirds appear, as shown in Figure 5.1.4. Actually, this complex has a particular structure that is interesting to understand: we can observe that it is in two components, and each of them is made up of major thirds for the vertical edges, and augmented fourths for the horizontal ones. We have chosen to give a second representation in Figure 5.1.5 of these components by drawing them into the chromatic circle: in fact, each component has exactly six pitch-classes, the "odd-ones" and the "even-ones", which are given by the following step of major seconds in the circle. Therefore, we have linked the pitches with major thirds and augmented fourths (and diagonals) to underline the symmetric disposition of the pitches in the complex.

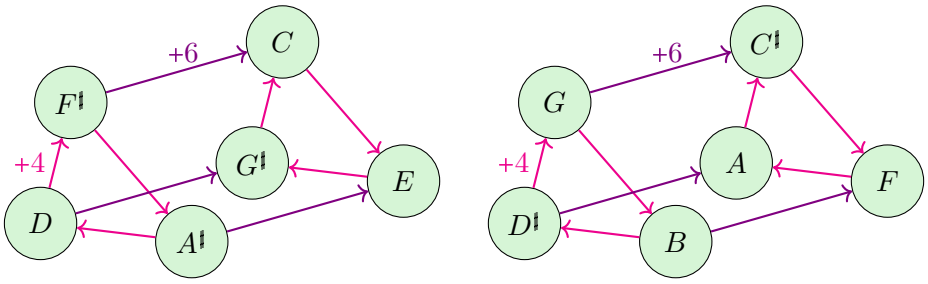


Figure 5.1.4: The associated complex with the chromatic scale and DFT distance at 91%: it is split in two connected components.

Furthermore, with a scaling parameter of 91%, the dimension of the homology group in degree 1 goes from 0 to 4. We have two generators for each component, which are given by alternating major thirds and augmented fourths, as shown in Figure 5.1.6. Since in our construction the triangles are filled, we can also extract a third one-dimensional cycle for each component.

Finally, the two components from Figure 5.1.4 are gathering at 95% of the filtration, and we show this in Figure 5.1.7: each pitch-class has now exactly five neighbors, given by the minor third (+3) and the reversed minor third (+9 = −3). Generators of homology in degree 1 from Figure 5.1.6 are still present in the complex, but we also have new ones given by alternating augmented fourth and reversed minor third. At this point of the filtration, we have a lot of edges and thus a lot of (filled) triangles, so many cycles in degree 1 are the same in H_1.

From 95% to 100%, the other harmonic intervals (±2, ±1 and ±5) appear successively in the filtration. This confirms the fact that the DFT sets the different harmonic intervals using the successive divisors of 12 (and their opposite in $\mathbb{Z}/12\mathbb{Z}$), as shown in Table 5.1.2 of distances.

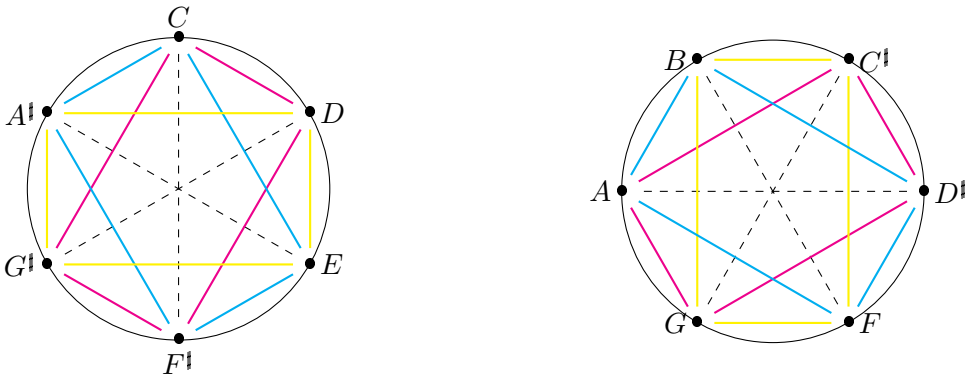


Figure 5.1.5: Another representation of the complex from Figure 5.1.4: each circle is a component, and the pitches are gathered by alternating major thirds and augmented fourths but without taking the order into account.

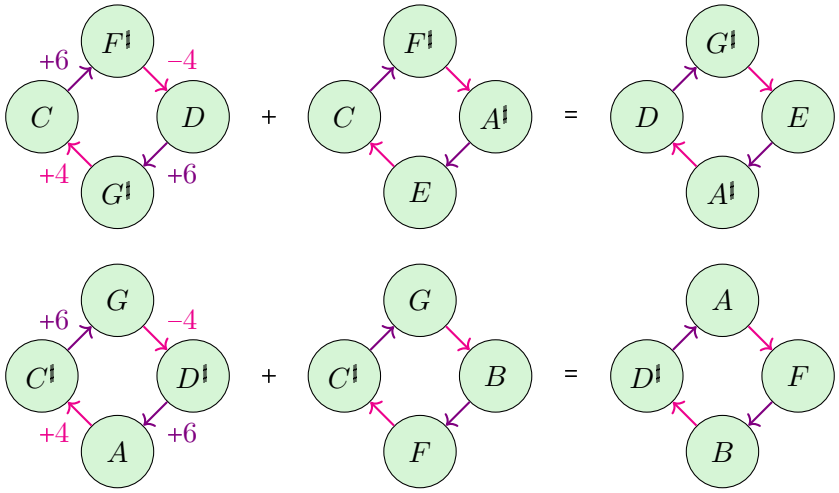


Figure 5.1.6: The one-dimensional cycles from the filtered complex, which appear at a scale of 91%: only two cycles generate H_1 for each component, and the third is a sum of the first two up to boundaries.

✔ **Conclusion for the chromatic scale.** We have seen that the metric given by the DFT applied to the set of all pitches gives a particular clustering, since a note is directly related to its augmented fourth (C with $F\sharp$). This first result seems musically questionable, but we will see that it becomes consistent when we choose our notes in some particular sets of pitches, such as more specific scales.

5.1.2. The diatonic scale

In order to explore this question of metrics on a set of notes, we could try to change the initial set of pitches we are working with: in fact, when we want to apply the same method to a set of chords, the natural approach would be to look first at a set of particular chords, such as minor and major chords (the Tonnetz $T[3,4,5]$). In the previous paragraph, we have chosen to look at the chromatic scale, which means all the pitches in $\mathbb{Z}/12\mathbb{Z}$ together, but we could also study some particular scales. We will start by looking at the major and minor diatonic scales.

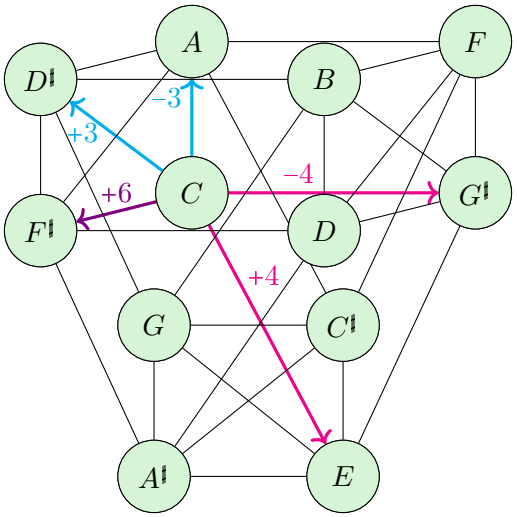


Figure 5.1.7: The filtered complex with a scaling parameter of 95%. Each point has exactly five neighbors: its minor and major third (+3, +4), their reversed (−3, −4) and its augmented fourth (+6).

❋ ***C*-Major scale.** Let us start with the major scale given by the fundamental C, which is shown in Figure 5.1.8. The set of pitches is also simply given by

$$\{C, D, E, F, G, A, B\} = \{0, 2, 4, 5, 7, 9, 11\}.$$

Following on from what we have done for the chromatic scale (Table 5.1.2), we can say that the closest pitch to C are successively the major and minor thirds E and A, the major second D and finally the fourth F, the fifth G and the seventh B. We have also computed the corresponding family of barcodes in degree 0 and 1, and the result is presented in Figure 5.1.9.

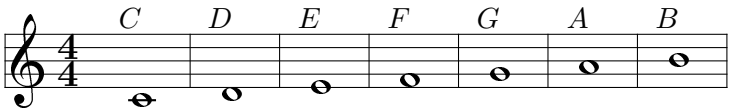


Figure 5.1.8: The C-major scale with whole notes in $\mathbb{Z}/12\mathbb{Z}$.

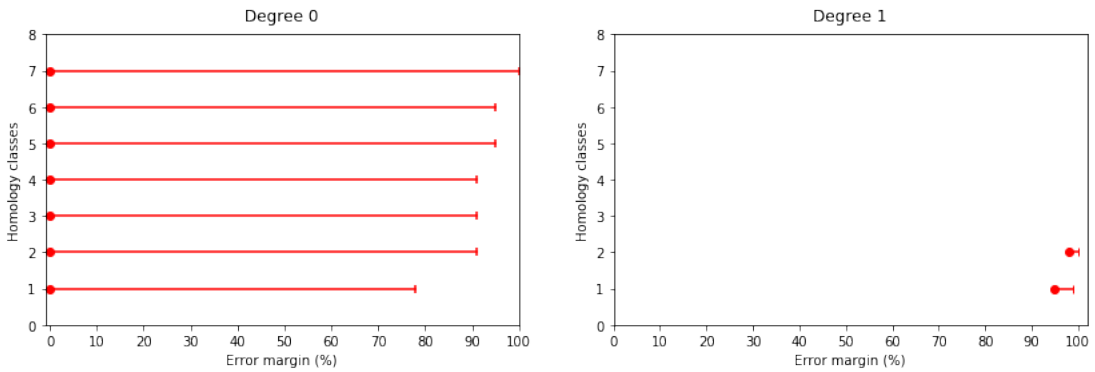


Figure 5.1.9: The associated family of barcodes with both C-major and C-minor scale.

We start by analyzing the barcode in degree 0: the first connection is given at 78%, and links together vertices F and B. In fact, these two pitches are the only ones in the C-major scale that are exactly separated by an augmented fourth. Then, there are only two filtration steps, at a scale of 91% and 95%, respectively. These two moments are illustrated in Figure 5.1.10. With a scaling parameter of 91%, the pitch C is immediately joined to its major third E, while the two other major thirds $\{G, B\}$ and $\{F, A\}$ are joined to the augmented fourth $\{B, F\}$.

At 95% of the filtration, the corresponding graph with the C-major scale takes on a remarkable shape: in fact, a circle of alternating major and minor thirds is formed:

$$\{C, E\} + \{E, G\} + \{G, B\} + \{B, D\} + \{D, F\} + \{F, A\} + \{A, C\}.$$

This circle is also the generator of homology in degree 1, that means the one-dimensional cycle that lasts from 95% to 99% of the filtration. Furthermore, this circle of thirds seems to be a nice illustration and characteristic topological feature of the major scale, and the fact that it appears by using the DFT as a metric to build the associated point cloud reinforces and confirms the fact that this approach has a particular musical meaning that needs to be exploited. This idea will also be confirmed when we will build the associated graph-type with a score in Section 6.4, and find back this particular circle of thirds for any major scale.

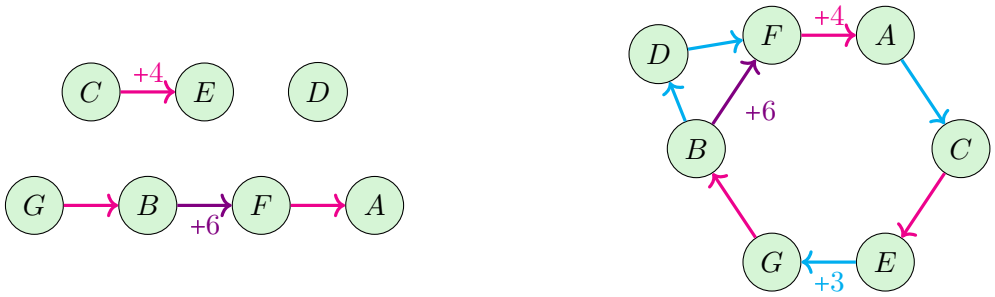


Figure 5.1.10: The associated filtration with the C-major scale with a scale of 91% (left) and 95% (right), respectively. The graph at 95% also represents the one-dimensional generator that lasts until 99% of the filtration, which is a circle of alternating major and minor thirds.

In degree 1, we see that there is also a second generator that starts only at 98% of the filtration but lasts until 100%. The associated complex is presented in Figure 5.1.11: we see that the major seconds appear on the graph and provide three new one-dimensional cycles, that are actually the same up to boundaries:

$$\{C, A\} + \{A, G\} + \{G, E\} + \{C, E\}$$

$$\{C, A\} + \{A, F\} + \{F, D\} + \{D, C\}$$

$$\{E, D\} + \{D, E\} + \{E, G\} + \{G, B\}$$

The circle of minor and major thirds is preserved in this configuration, and since these new one-dimensional cycles are shorter than the first one, we might then consider the circle of thirds to be most representative of the major scale.

✻ C-minor scale. Let us do the same analysis for a minor scale. For comparison, we naturally choose to study the C-minor scale (Figure 5.1.12), with associated set of pitches

$$\{C, D, D\flat, F, G, G\flat, A\flat\} = \{0, 2, 3, 5, 7, 8, 10\}.$$

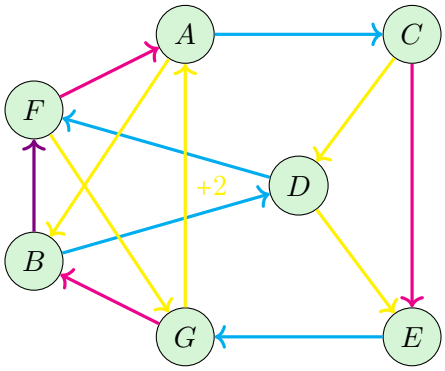


Figure 5.1.11: The associated filtration with the C-major scale at a scale of 98%. The new edges are given by the major seconds $\{C, D\}$, $\{D, E\}$, $\{G, A\}$ and $\{A, B\}$, which provide a new one-dimensional cycle that lasts until 100% of the filtration.

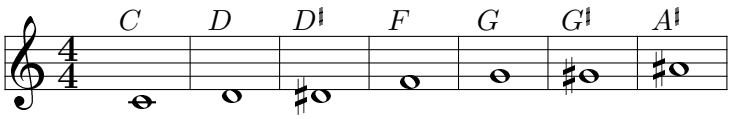


Figure 5.1.12: The C-minor scale with whole notes in $\mathbb{Z}/12\mathbb{Z}$.

We can see that the major and minor scales are connected by the minor third transposition $z \mapsto z + 3$ for $z \in \mathbb{Z}/12\mathbb{Z}$, which sends the C-major scale (C, D, E, F, G, A, B) to the $D♯$-major scale $\left(D♯, F, G, G♯, A♯, C, D\right)$. This transposition corresponds to a rotation, and Theorem 2.4.5 from Chapter 2 tells us that such transformations are isometries for the DFT-distance. Therefore, the resulting barcodes and filtrations will be the same for major and minor scales (with just a relabelling of the vertices). In particular, we will recover the circle of thirds from Figure 5.1.10 as a characteristic shape of the C-minor scale, according to the third transposition of pitches.

5.1.3. The pentatonic scale

In this paragraph, we simply replace the diatonic scales with the pentatonic ones, and we keep C as our fundamental pitch. Let us start with the major pentatonic scale, which is shown in Figure 5.1.13. The set of pitches is given by

$$\{C, D, E, G, A\} = \{0, 2, 4, 7, 9\}.$$

Following on from what we have done for the chromatic scale (Table 5.1.2), we can say that the pitches that are closest to C are successfully the major and minor thirds E and A, the major second D and finally the fifth G.

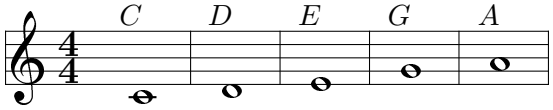


Figure 5.1.13: The C-major pentatonic scale with whole notes in $\mathbb{Z}/12\mathbb{Z}$.

We have computed the associated family of barcodes in degree 0 and 1, and the results are presented in Figure 5.1.14. Recall that in this particular example, we only have five musical bars, so we do not expect complex barcodes. With a scaling parameter of 91%, the major third

$\{C, E\}$ appears on the filtration, and, at 95%, the two minor thirds $\{C, A\}$ and $\{E, G\}$ follow. As for the general major scale, the pitch D is still an isolated vertex at this point.

With a scaling parameter of 98%, the filtration is thus in one connected component, which is given by the connection of all the major seconds $\{C, D\}$, $\{D, E\}$ and $\{G, A\}$. Moreover, at this moment there is a one-dimensional cycle, visible on the barcode in degree 1, which remains only one time of the filtration. This cycle is given by two minor thirds, one major and one second ones:

$$\{C, E\} + \{E, G\} + \{G, A\} + \{A, C\}.$$

The graphs with a scaling parameter of 95% and 98% are shown in Figure 5.1.15, where the one-dimensional cycle is visible. Furthermore, this particular cycle is also the complementary of the sequence of fifths on which the major pentatonic scale is built:

$$\{C, G\} + \{G, D\} + \{D, A\} + \{A, E\}.$$

This complementary relationship is represented in Figure 5.1.16. The fourth $\{E, A\}$ appears at 99% and the graph is fully connected at 100%.

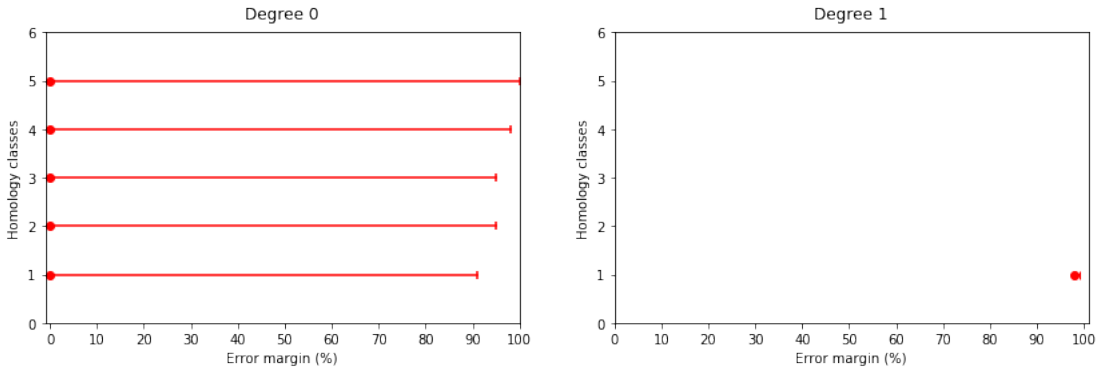


Figure 5.1.14: The associated family of barcodes with both C-major and C-minor pentatonic scales.

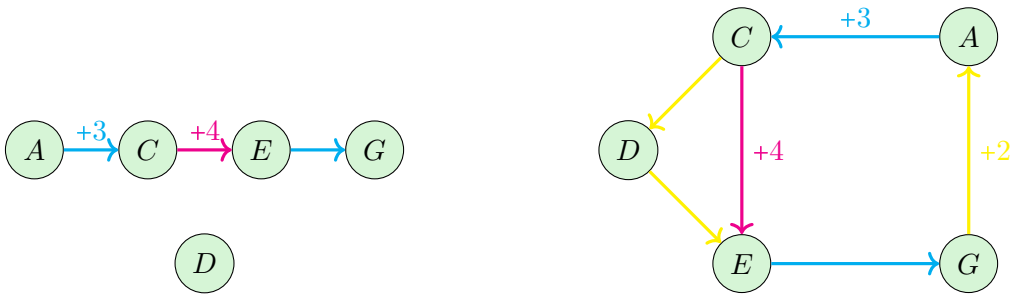


Figure 5.1.15: The associated filtration with the C-major pentatoic scale with a scale of 95% (left) and 98% (right), respectively. The graph at 98% represents also the only one-dimensional generator that lasts until 99% of the filtration.

Finally, as for the case of the diatonic scales, the minor pentatonic scale is given by the major one with a translation of pitches $z \mapsto z + 3$ for $z \in \mathbb{Z}/12\mathbb{Z}$, which sends (C, D, E, G, A) to $\left(D\sharp, F, G, A\sharp, C\right)$. Therefore, such rotations of pitches are isometries (Theorem 2.4.5 from Chapter 2), so the resulting barcodes and filtrations will be the same for pentatonic major and minor scales (with just a relabelling of the vertices). In particular, we will recover the characteristic shapes from Figures 5.1.15 and 5.1.16 for the C-minor pentatonic scale.

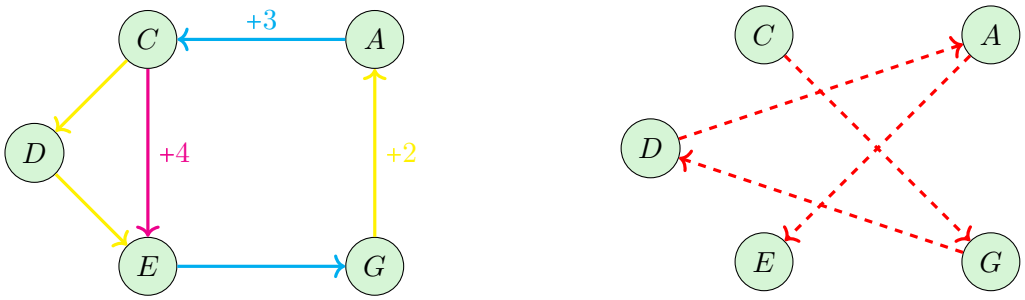


Figure 5.1.16: The associated filtration with the C-major pentatonic scale, with a scaling parameter respectively of 98% (left) and its complement (right): we recover the sequence of fifths on which the pentatonic major scale is built.

✔ **Conclusion for the diatonic and pentatonic scales.** These last two examples provide rewarding examples of the application of the DFT as a metric on some specific set of pitches. In particular, it shows that the classification and comparison of notes depend on the choice of the initial set. It also allows us to highlight some interesting topological features that we can associate with different scales, which we will recover in Section 6.4. Our aim now is to generalize this construction to compare chords instead of single notes.

5.2. A METRIC ON THE SET OF 3-CHORDS

The aim of this section is to apply the metric given by the DFT to some particular sets of chords. More precisely, we will assume that a chord is given by **three notes**, so we will really be talking about 3-chords. In the previous paragraph, we first decided to use the DFT as a metric between all the pitches in $\mathbb{Z}/12\mathbb{Z}$, and the results were inconclusive. Therefore, we adjusted the initial set of pitches by taking certain scales and then found much more interesting topological features.

In this purpose, we will choose the set of 3-chords to work with: a natural start might be to take all the major and minor chords, that means the twenty-four elements of the Euler's Tonnetz. Then we can generalize to all the two-dimensional Tonnetze of the classical form $T[a,b,c]$. Before doing so, we will begin this section with a rigorous definition of the two-dimensional Tonnetz.

5.2.1. THE TWO-DIMENSIONAL TONNETZ: DEFINITION

In the literature, the Tonnetz is presented under many different forms, such as a simplicial complex, a lattice or a topological space (see for example [15] or [33]). Our aim here is to formalize these definitions in order to be able to use the word "Tonnetz" without any ambiguity.

Definition 5.2.1. Let $a,b,c \in \mathbb{Z}/12\mathbb{Z}\setminus\{0\}$ be such that $a+b+c=0$. We define a simplicial complex $T[a,b,c]$ of dimension two with $\mathbb{Z}/12\mathbb{Z}$ as its set of vertices, and where

- $\{x,y\}$ is an edge if and only if $x-y \in \{\pm a, \pm b, \pm c\}$
- the triangles are the sets of the form $\{x, x+a, x+a+b\}$ or $\{x, x+b, x+a+b\}$

For simplicity, the notation $T[a,b,c]$ may refer in what follows to the corresponding topological space $|T[a,b,c]|$.

Remark 5.2.2. Permuting a,b,c does not affect $T[a,b,c]$. For example, $T[c,b,a]$ contains triangles of the form $\{x, x+c, x+c+b\}$: then, by putting $y = x+c$, this set becomes $\{y, y+b, y+a+b\}$, which is a triangle of $T[a,b,c]$. Similarly, the simplicial complex $T[-a,-b,-c]$ contains the same triangles as $T[a,b,c]$, so we have $T[a,b,c] = T[-a,-b,-c]$.

From this last remark, we can enumerate all the two-dimensional Tonnetz, and there are twelve of them:

$$T[1,1,10],\ T[1,2,9],\ T[1,3,8],\ T[1,4,7],\ T[1,5,6],\ T[2,2,8]$$
$$T[2,3,7],\ T[2,4,6],\ T[2,5,5],\ T[3,3,6],\ T[3,4,5],\ T[4,4,4]$$

Furthermore, with the work done in [33], each of these Tonnetze can be described by direct inspection in a topological way, as shown in Table 5.2.1. We will explain below how the different topologies appear from Definition 5.2.1.

Tonnetz	Topology	Betti number		
		β_0	β_1	β_2
$T[1,2,9]$, $T[2,3,7]$ $T[1,3,8]$, $T[3,4,5]$ $T[1,4,7]$	Torus	1	2	1
$T[1,1,10]$, $T[2,5,5]$	Cylinders	1	1	0
$T[1,5,6]$	Necklace of six tetrahedra	1	1	6
$T[2,2,8]$	Two cylinders	2	2	0
$T[2,4,6]$	Two necklaces of three tetrahedra	2	2	6
$T[3,3,6]$	Three tetrahedra	3	0	3
$T[4,4,4]$	Four triangles	4	0	0

Table 5.2.1: Classification of the twelve two-dimensional Tonnetze according to their topology.

We now want to give a visual representation of a Tonnetz, so we are going to represent it as a lattice. More precisely, if $\{e_1, e_2\}$ is the canonical basis of $\mathbb{Z}^2$, let us consider the homomorphism $\phi = \phi_{a,b,c} : \mathbb{Z}^2 \to \mathbb{Z}/12\mathbb{Z}$ defined by

$$\phi(e_1) = a, \quad \phi(e_2) = b$$

so that $c = -\phi(e_1 + e_2)$. It is easy to see that the image of ϕ is the connected component of 0 in $T[a,b,c]$. More generally, $\mathbb{Z}/12\mathbb{Z}$ is given by the disjoint union of the connected components of $T[a,b,c]$, so their number is given by the index of the image of ϕ in $\mathbb{Z}/12\mathbb{Z}$.

We take now $\{\epsilon_1, \epsilon_2\}$ a basis of $\mathbb{R}^2$, and we draw the lattice L generated by these, which we identify with $\mathbb{Z}^2$ using the basis just chosen. Then, we define a structure of infinite simplicial complex on $\mathbb{R}^2$ in a way that is similar to Definition 5.2.1, i.e. the vertices are the elements of L and

- $\{x, y\}$ is an edge when $x - y \in \{\pm\epsilon_1, \pm\epsilon_2, \pm(\epsilon_1 + \epsilon_2)\}$
- the triangles are the sets of the form $\{x, x + \epsilon_1, x + \epsilon_1 + \epsilon_2\}$ or $\{x, x + \epsilon_2, x + \epsilon_1 + \epsilon_2\}$

Thus, we can see our homomorphism ϕ as a map between the set of vertices of $\mathbb{R}^2$ and $T[a,b,c]$. By construction, it is a simplicial map. The geometric realization $|L|$ can be seen as a triangulation of $\mathbb{R}^2$, and we then obtain a continuous map $|\phi| : \mathbb{R}^2 \to T[a,b,c]$ (see Definition 3.1.2 from Chapter 3). The image of $|\phi|$ is the connected component containing 0, given our conventions, and this gives us a representation of $T[a,b,c]$ as a lattice.

To make it even more visual, next to each lattice point $v \in L$, we write $\phi(v)$ as a label, usually using the musical dictionary $0 = C, 1 = C\sharp$ and so on. The picture finally obtained is often colloquially refered to as the Tonnetz itself (the word "Tonnetz" being German for "network of notes", see [41]). For instance, we give the visual representation of $T[3,4,5]$ and $T[4,4,4]$ in Figure 5.2.1.

In Table 5.2.1, we speak about the topology of a Tonnetz $T[a,b,c]$. To visualize this, it is helpful to define the fundamental domain of a Tonnetz. Let us start by observe that the map $|\phi|$ is far from being injective, and in fact it can make drastic identifications. Firstly, if we consider $\ker\phi$ as a subgroup of L and thus of $\mathbb{R}^2$, we have

$$|\phi|(x+k) = |\phi|(x)$$

for all $k \in \ker\phi$. In fact, the quotient of $\mathbb{Z}^2$ by $\ker\phi$ is a subgroup of $\mathbb{Z}/12\mathbb{Z}$, so $\ker\phi$ is a finite-index subgroup of L, and thus is a lattice itself generated say by u_1 and u_2. In the case of $T[3,4,5]$, we can take $u_1 = 3\epsilon_1$ and $u_2 = 4\epsilon_2$. We write $D_{a,b,c}$ for the parallelogram with the origin as a vertex and u_1 and u_2 as sides, and we call $D_{a,b,c}$ the **fundamental domain** for the Tonnetz $T[a,b,c]$ (with respect to the choice of u_1 and u_2, which are not unique). We give the representation for $T[3,4,5]$ and $T[4,4,4]$ in Figure 5.2.1a and 5.2.1b respectively. Moreover, we have $|\phi|(D_{a,b,c}) = T[a,b,c]$ when ϕ is surjective (otherwise $|\phi|(D_{a,b,c})$ is simply the connected component containing 0) so, in other words, the Tonnetz $T[a,b,c]$ is obtained from the fundamental domain by making identifications. More precisely, let us define

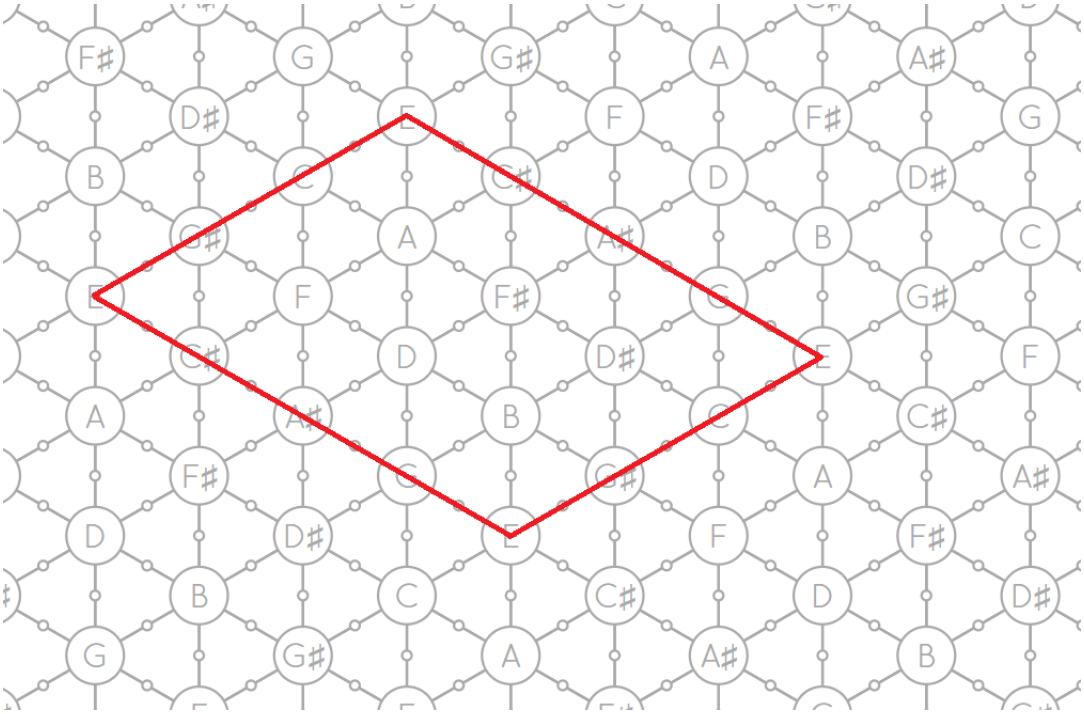
$$\mathbb{T}_{a,b,c} := \mathbb{R}^2 / \ker\phi$$

so that $\mathbb{T}_{a,b,c}$ is a torus obtained from $D_{a,b,c}$ by identifying the opposite sides. In fact, there is an induced map $\mathbb{T}_{a,b,c} \to T[a,b,c]$ and, in some cases, such as that of $T[3,4,5]$, it is a homeomorphism. On the other hand, for $T[4,4,4]$, every single triangle of $\mathbb{R}^2$ is sent by ϕ to the unique triangle in the corresponding connected component of $T[4,4,4]$, which is consistent with the classification Table 5.2.1 already given.

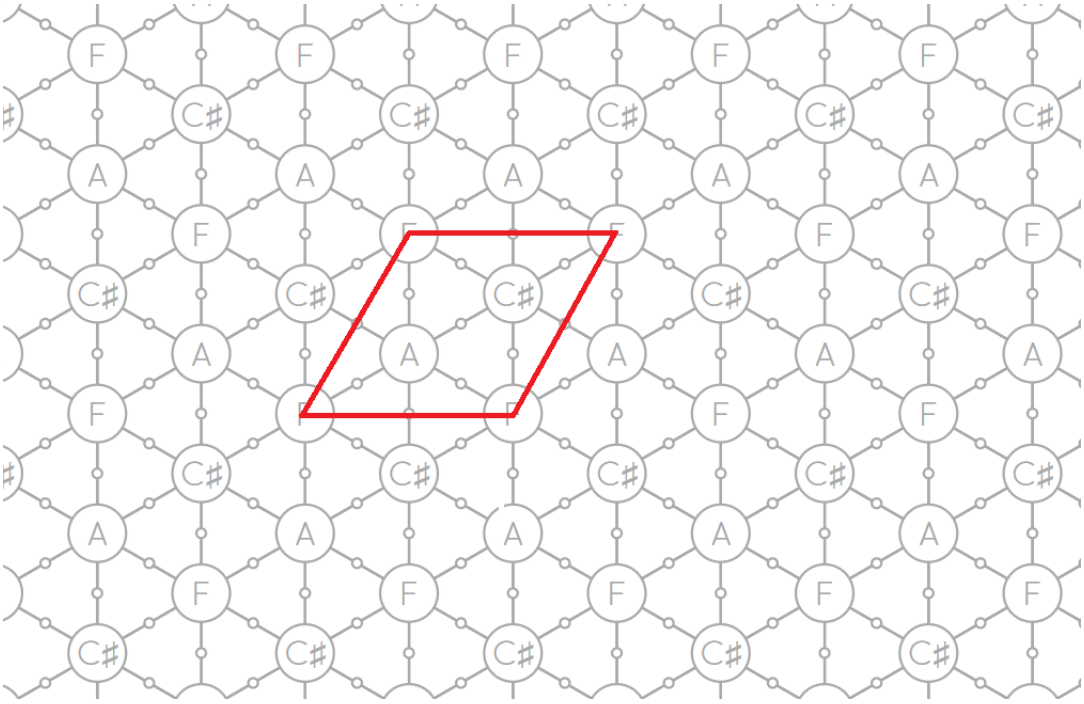
Remark 5.2.3. We can go a little further with the classifications of the Tonnetze: in fact, still according to [33], we can arrange them in pairs by considering basic transformations. More precisely, there is an isomorphism of simplicial complex between $T[3,4,5]$ and $T[1,3,8]$, which is given by the homothetie $z \mapsto 5z$ for $z \in \mathbb{Z}/12\mathbb{Z}$. Therefore, using Theorem 2.4.14 from Chapter 2, this map is an isometry so the filtrations and barcodes for $T[1,3,8]$ and $T[3,4,5]$ will be exactly the same. The same identification holds for the pair of Tonnetze $T[1,2,9]$ and $T[2,3,7]$ ($z \mapsto 7z$), and also for $T[1,1,10]$ and $T[2,5,5]$ ($z \mapsto 5z$).

5.2.2. The Tonnetz of major and minor chords

Let us focus on the Tonnetz $T[3,4,5]$. Musically, this Tonnetz is the one given by minor and major chords: indeed, if we recall the fundamental domain $D_{3,4,5}$, one side is given by the major thirds (+4), while the other is given by the minor thirds (+3), so the diagonals are given by the fifths (+7 = −5). Actually, this is the most famous Tonnetz because it is the original one, defined by Leonhard Euler (1707-1783) himself in 1739 for historical musical reasons: his idea at that time was to group pitch-classes according to their acoustic proximity (see [41]). For this purpose, he only used majord thirds and fifths, so a chord was not correctly represented as a triangle. The Tonnetz was then generalized by Arthur von Oettingen (1836-1920) and Hugo Riemann (1849-1919) who added minor thirds and extended the Tonnetz as an infinite plan. Such construction is sometimes called the **Neo-Riemannian Tonnetz** (see [42]). The eleven other two-dimensional Tonnetze $T[a,b,c]$ are an extension of this definition, coming much later with [15], [21] or [33] for example.



(a) The Tonnetz $T[3,4,5]$ as a lattice with its fundamental domain $D_{3,4,5}$.



(b) The Tonnetz $T[4,4,4]$ as a lattice with its fundamental domain $D_{4,4,4}$.

Figure 5.2.1: Visual representation of the Tonnetze $T[3,4,5]$ and $T[4,4,4]$ as lattices together with their respective fundamental domain $D_{3,4,5}$ and $D_{4,4,4}$. The pictures are extracted from the web-hexachord application `https://guichaoua.gitlab.io/web-hexachord/`.

Our purpose here is to apply the DFT together with persistent homology to the set of major and minor chords: as we did in the previous Section 5.1, we will associate a score with the Tonnetz (instead of a scale) and then compute persistent homology using the DFT on this point cloud.

We will start by analyzing the barcode in degree 0 obtained from the filtration built on the set of major and minor chords, and it will turn out that the PLR-group of basic transformations between such chords will emerge from the study. Let us begin with a few reminders about this group. Then, the barcode in degree 1 will provide some interesting one-dimensional cycles, which we will describe manually.

★ The PLR-group in degree 0.

A Tonnetz is a simplicial complex of dimension two, so we can define some applications that transform a triangle from the Tonnetz into another. In the particular case of $T[3,4,5]$, a triangle corresponds to either a major or a minor chord and it happens that, for a given triangle let us say a major one, the three neighbors are given by a minor one. Therefore we define three basic transformations that change a minor chord into a major one and vice versa. These operations are sometimes called the **Neo-Riemannien transformations** (defined by Hugo Riemann in 1880) and are denoted by:

$$P \text{ (\textbf{parallel})}, \ L \text{ (\textbf{leading-tone})} \text{ and } R \text{ (\textbf{Relative})}.$$

We start by giving some basic definitions and properties of these operations that we will use in the rest of this section, and we refer to [42] for more details on the subject.

Definition 5.2.4. A **triad** is a chord made up of three notes (pitch-classes): a **root** r, a **third** (minor t_m or major t_M) and a **fifth** f, with r, t and f in $\mathbb{Z}/12\mathbb{Z}$. Obviously, the choice of t and f depends on the root, and we have $t_m = r+3$, $t_M = r+4$ and $f = r+7$. We denote by $r_M = \{r, t_M, f\}$ and $r_m = \{r, t_m, f\}$ respectively the major and minor chord built from the root r. By successively fixing r, t_M (or t_m) and f, we can transform r_M (or r_m) into a minor (or major) chord with the fewest possible moves (fixing one edge for each transformation):

- $P(r_M) = \{r, t-1, f\} = r_m$
- $L(r_M) = \{r-1, t, f\} = (r+4)_m$
- $R(r_M) = \{r, t, f+2\} = (r-3)_m$
- $P(r_m) = \{r, t+1, f\} = r_M$
- $L(r_m) = \{r, t, f+1\} = (r-4)_M$
- $R(r_m) = \{r-2, t, f\} = (r+3)_M$

These three transformations P, L and R are given by basic reflections on the chromatic circle: if we take $C = 0_M = \{0,4,7\}$ as an example, then we get that

$$\begin{array}{rcccl} P(C) & = & 0_m & = & \{0,3,7\} & = & Cm \\ L(C) & = & 3_m & = & \{11,4,7\} & = & Em \\ R(C) & = & 9_m & = & \{0,4,9\} & = & Am \end{array}$$

An illustration of these operations is given in Figure 5.2.2. It is important here to notice that the resulting minor chords are the exact three neighbors of C in the Tonnetz $T[3,4,5]$.

Moreover, these three operations are involutions ($P^2 = L^2 = R^2 = 1$) and form a group called the PLR-**group**, which is generated by L and R with $P = R(LR)^3$. The element $(RP)^2 = (PR)^2$ is sometimes simply denoted by T_6: in fact, for a chord given by root $r = \{r, t, f\}$ (major or minor), T_6 is given by the translation of an augmented fourth

$$T_6(r) = \{r+6, t+6, f+6\},$$

which is equivalent to applying the operation $(RP)^2 = (PR)^2$. Finally, we have the following proposition from [42]:

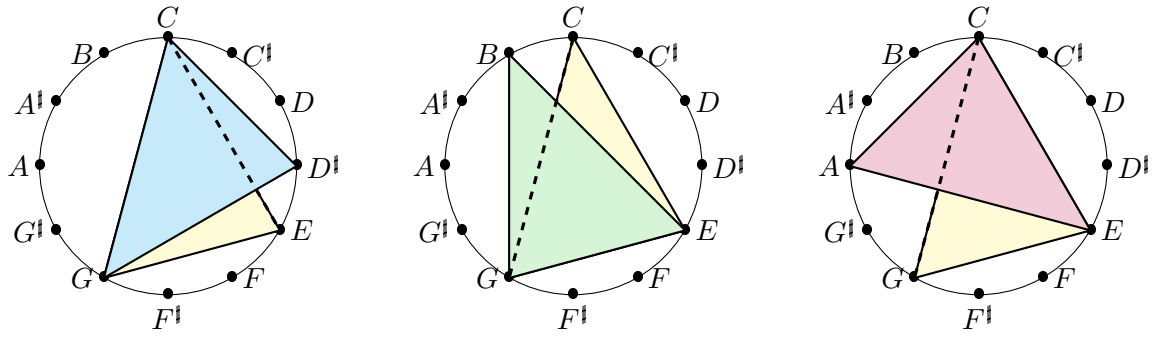


Figure 5.2.2: P, L and R transformations on the major chord $C = \{0, 4, 7\}$:

$$P(C) = Cm,\ L(C) = Em \text{ and } R(C) = Am$$

Proposition 5.2.5. *The action of the PLR-group on the set of the twenty-four minor and major chords is simply transitive.*

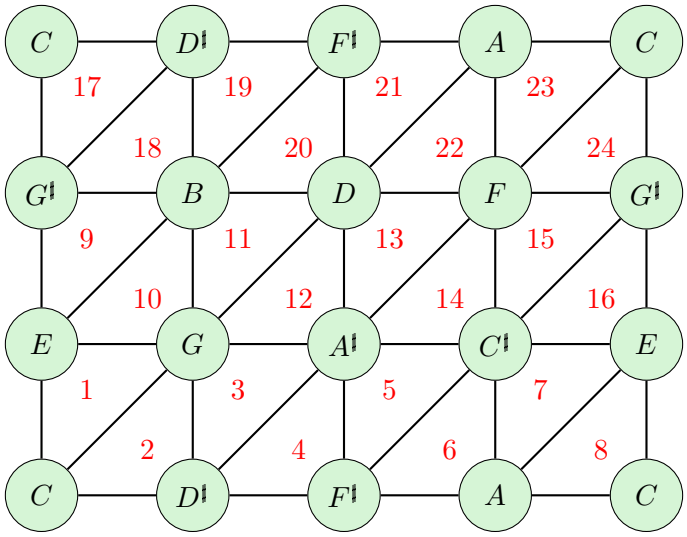
Let us go back to our model: there are exactly twenty-four chords in the fundamental domain of $T[3, 4, 5]$, so we can set a bijection between $D_{3,4,5}$ and $\{1, 2, \ldots, 24\}$, as we did in Figure 5.2.3, where we start with the C-major chord as $\mathcal{B}_1$. Therefore, there is a corresponding score $\mathcal{S}_{T[3,4,5]} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{24}\}$ with the Tonnetz $T[3, 4, 5]$, also shown in Figure 5.2.3.

As we did for the scales in Equation 5.1.1, we can directly compute the distance between two chords using the Fourier coefficients: more precisely, if r and r' are two (minor or major) chords given by the set of pitches $\{r, t, f\}$ and $\{r', t', f'\}$ respectively, the distance between r and r' is given by

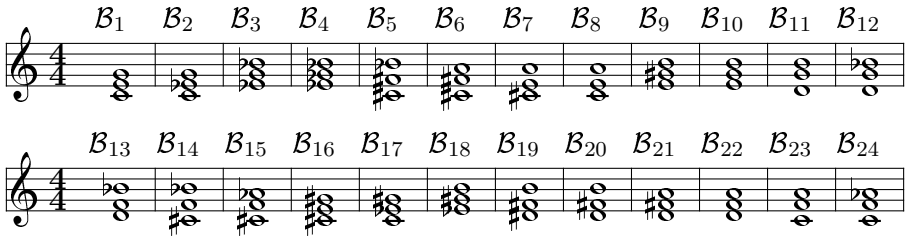
$$\begin{aligned} \mathrm{d}_{\mathrm{DFT}}(r, r') &= \sum_{y=1}^{12} \left| \mathcal{F}_{r,t,f}(y) - \mathcal{F}_{r',t',f'}(y) \right| \\ &= \sum_{y=1}^{12} \left| \left(\exp\left(\frac{-2i\pi ry}{12}\right) + \exp\left(\frac{-2i\pi ty}{12}\right) + \exp\left(\frac{-2i\pi fy}{12}\right)\right) \right. \\ &\quad \left. - \left(\exp\left(\frac{-2i\pi r'y}{12}\right) + \exp\left(\frac{-2i\pi t'y}{12}\right) + \exp\left(\frac{-2i\pi f'y}{12}\right)\right) \right| \end{aligned} \tag{5.2.1}$$

Therefore, we can compute the distances between each chord of the Tonnetz $T[3, 4, 5]$ by simply using the above calculation. We summarize all this information in Table 5.2.2, where we have chosen to include the correspondence between a chord (C as an example) and operations of the PLR-group, whose action on the Tonnetz is simply transitive (Proposition 5.2.5). Finally, from this table we can compute the associated family of barcodes, which is shown in Figure 5.2.4.

Let us analyze these two barcodes in degree 0: there are two characterized moments, successive in time, at 58% and 59% of the filtration. At a scale of 58%, the complex looks like in Figure 5.2.5, where there are exactly twelve connected components. Notice that the chords are grouped together with their relative ones. For instance, C and Am are connected by an edge. In fact, using 5.2.1, the distance between $C = \{0, 4, 7\}$ and $Am = \{9, 0, 4\}$ is simply given by the following calculation:



(a) The fundamental domain $D_{3,4,5}$ of the Tonnetz $T[3,4,5]$ with an arbitrarily numbered chords.



(b) The score $\mathcal{S}_{T[3,4,5]} = \{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_{24}\}$ made up of the twenty-four minor and major chords of the Euler's Tonnetz.

Figure 5.2.3: The associated score $\mathcal{S}_{T[3,4,5]}$ with the fundamental domain $D_{3,4,5}$ of the Euler's Tonnetz.

id	R	P	L	$PLR = RLP$	PL	LP	T_6
0	14.928	15.192		17.856	17.949		18.742
C	Am	Cm	Em	Fm	E	$G♯$	$F♯$

$PRL = LRP$	$RPL = LPR$	RP	PR	RL	LR	RLR	LRL
18.928		19.253		20.325		21.413	21.856
Gm	$C♯m$	$D♯$	A	G	F	Dm	Bm

RT_6	PLT_6	LPT_6	RLT_6	LRT_6	PT_6	LT_6	LPL
22.928	23.307		24.325		24.435		25.555
$D♯m$	$A♯$	D	$F♯m$	$A♯m$	$C♯$	B	$G♯m$

Table 5.2.2: The table of distances between the twenty-four major and minor chords (starting from the C-major chord): here we give the correspondence with the twenty-four elements of the PLR-group.

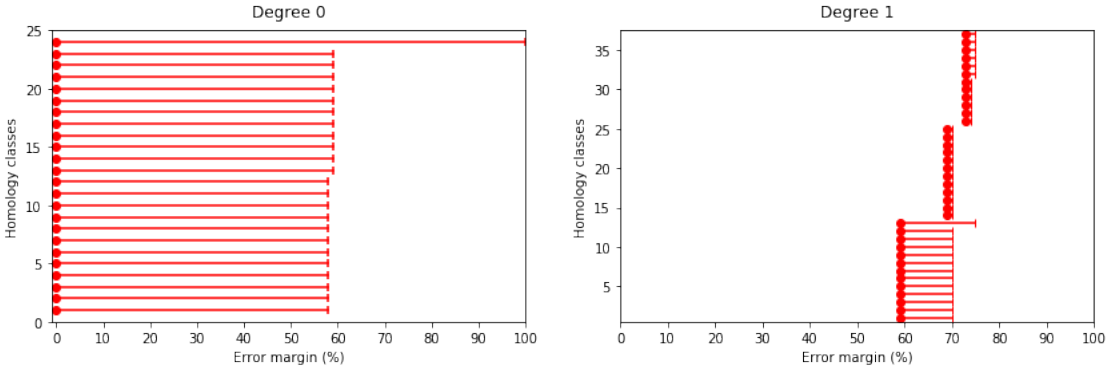


Figure 5.2.4: The associated family of barcodes with the Tonnetz $T[3,4,5]$.

$$\begin{aligned}
\mathrm{d}_{\mathrm{DFT}}(C, Am) &= \sum_{y=1}^{12} \left|\exp\left(\tfrac{-2\times 0 i\pi y}{12}\right) - \exp\left(\tfrac{-2\times 9 i\pi y}{12}\right)\right| \\
&= \sum_{y=1}^{12} |1 - i^y| \\
&= 3\big(|1-i| + |1+i| + 2\big) \\
&= 6\sqrt{2} + 6 \\
&\cong 14.485
\end{aligned}$$

Notice that this calculation is the same as computing the distance between the pitches C and A. In fact, just take the three neighbors of the C-major chord which are given by Em, Am and Cm, remove the two common notes and simply compute the distance between the remaining pitches, according to Table 5.1.2 from the previous section. In this case, we have

$$14.485 \cong \underbrace{\mathrm{d}_{\mathrm{DFT}}(C, A)}_{=\mathrm{d}_{\mathrm{DFT}}(C,Am)} \leq \underbrace{\mathrm{d}_{\mathrm{DFT}}(C, B)}_{=\mathrm{d}_{\mathrm{DFT}}(C,Em)} = \underbrace{\mathrm{d}_{\mathrm{DFT}}(C, D^\sharp)}_{=\mathrm{d}_{\mathrm{DFT}}(C,Cm)} \cong 15.192.$$

Therefore, the chord that minimizes the DFT-distance is given by its relative one, which is musically consistent.

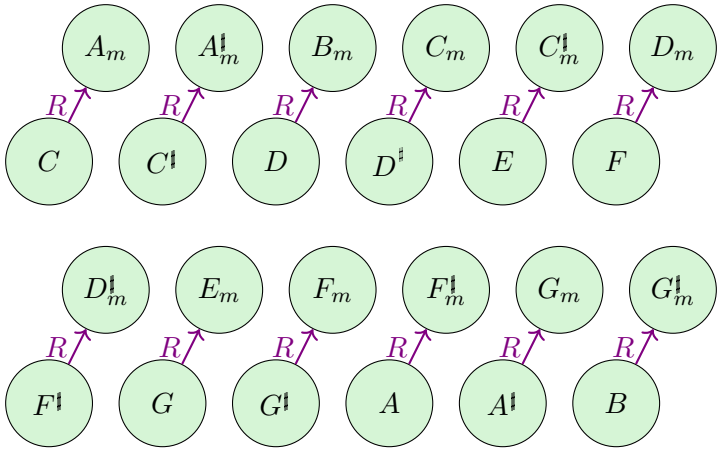


Figure 5.2.5: The graph of the filtration associated with the Tonnetz $T[3,4,5]$ at a scale of 58%.

With a scaling parameter of 59%, the complex looks like in Figure 5.2.6 (left), where we have shown only the graphs (with vertices in $\{1, 2, \ldots, 24\}$ from Tonnetz 5.2.3): in this illustration, we have chosen to show the complexes at 59% and 69% because the filtration stagnates at 59% (the first moment where it is connected) and remains the same until it becomes as in 69%. There are several things to notice here: firstly, in both cases we recover the shape of a torus, the quotiented fundamental domain of $T[3, 4, 5]$, with a relationship of duality since a triangle is a vertex in our graph. We also get some more informations considering the number of one-dimensional cycles that we will study in the next paragraph. Secondly, each vertex of the complex at 59% has exactly three neighbors: for instance, the C-major chord (bar $\mathcal{B}_1$, vertex $\{1\}$) is connected to the bars $\mathcal{B}_2$, $\mathcal{B}_8$ and $\mathcal{B}_{10}$. Looking back at the Tonnetz 5.2.3, we see that it corresponds respectively to the minor chords Cm, Am and Em, the three neighbors of C in the fundamental domain of the Tonnetz. In other words, the C-major chord is connected to its parallel, leading-tone and relative chord (Definition 5.2.4), which are exactly the three minor chords that minimize the DFT-distance from one given chord. This observation is illustrated in Figure 5.2.7 and also leads to Theorem 5.2.6.

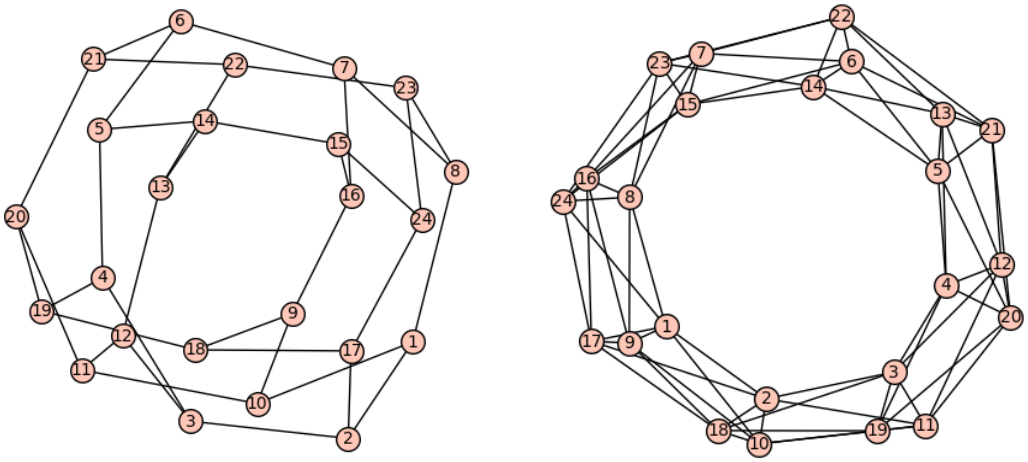


Figure 5.2.6: The filtration associated with the Tonnetz $T[3, 4, 5]$ at a scale of 59% (left) and 69% (right).

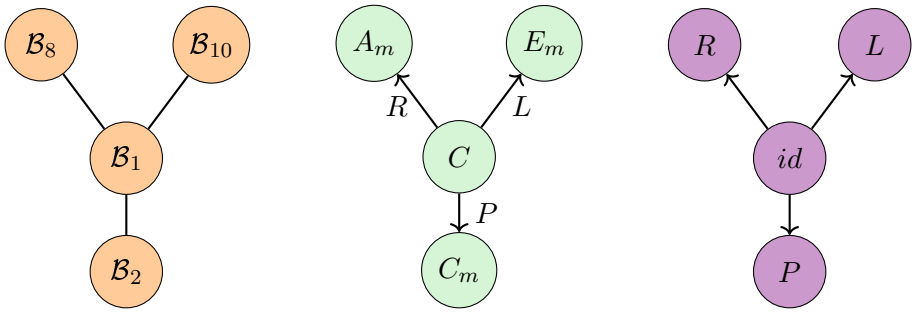


Figure 5.2.7: A zoom on the graph associated with the Tonnetz $T[3, 4, 5]$ at 59% of the filtration: each chord has exactly three neighbors that are given by the three basic transformations P, L and R. Here we give the example of the C-major chord with its neighbors $P(C) = Cm$, $L(C) = Em$ and $R(C) = Am$.

Theorem 5.2.6. *The graph associated with the twenty-four major and minor chords of the Tonnetz $T[3, 4, 5]$, given by the filtration below with a scaling parameter of 59%, corresponds to the Cayley graph of the PLR-group generated by the three transformations P, L and R.*

Proof. From Proposition 5.2.5, the action of the PLR-group on the set of the twenty-four major and minor chords is simply transitive, so there is a bijection between the elements of the PLR-group and the chords of the Tonnetz. Therefore, using the illustration from Figure 5.2.7, the left graph from Figure 5.2.6 is exactly the corresponding Cayley graph. ∎

★ One-dimensional cycles on the Tonnetz.

We can now focus on the barcode in degree 1 from 5.2.4: first, we observe that when the complex is fully connected (at a scale of 59%), there are twelve generators of homology H_1 rising, and one lasts until 74% of the filtration. By computing these classes by hand, we see that there are exactly three different types of cycles that can be described by the operations P, L and R:

✻ **First type of cycle.** Starting from a given note, take all the minor and major chords that contain it (see Figure 5.2.8): this corresponds to applying the operation $(PLR)^2$ to the corresponding major chord. For instance, starting from G, we get

$$(PLR)^2(G) = (PLR)PL(Em) = (PLR)P(C) = (PLR)(Cm) = (PL)(D♯) = P(Gm) = G$$

and we then obtain the cycle of length 6

$$\{G, Em\} + \{Em, C\} + \{C, Cm\} + \{Cm, D♯\} + \{D♯, Gm\} + \{Gm, G\}$$

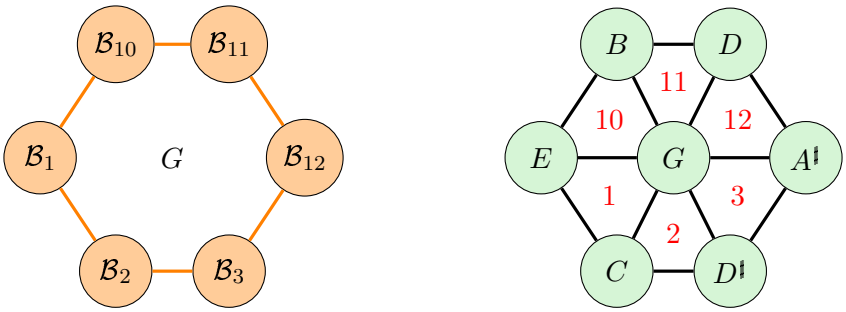


Figure 5.2.8: The homology class in degree 1 for $T[3,4,5]$ which is obtained by turning around one given note (here G) and taking all minor and major chords containing it. There are exactly six such chords for one given note so the result is a cycle of length 6.

✻ **Second type of cycle.** Starting from a minor third, take all the successive chords given by the following major third side (see Figure 5.2.9): this corresponds to applying the operation $(LP)^3$ to a starting chord. For instance, starting from the C-minor chord, we get

$$(LP)^3(Cm) = (LP)^2L(C) = (LP)^2(Em) = LPL(E) = LP(A♭m) = L(A♭) = C_m$$

and we then obtain the cycle of length 6

$$\{Cm, C\} + \{C, Em\} + \{Em, E\} + \{E, A♭m\} + \{A♭m, A♭\} + \{A♭, Cm\}$$

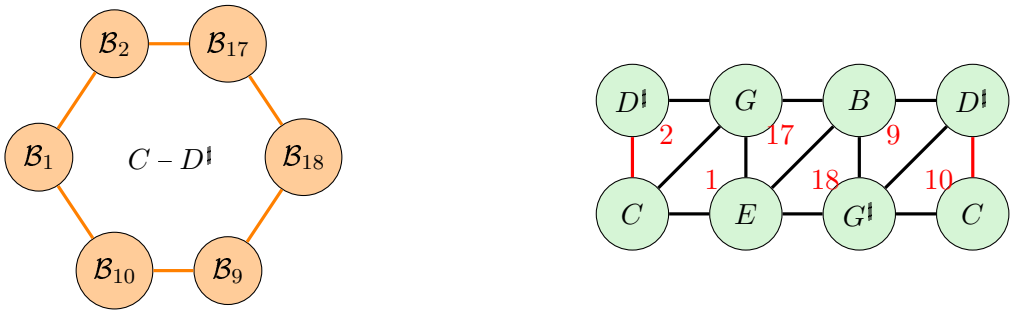


Figure 5.2.9: The homology class in degree 1 for $T[3,4,5]$, obtained by starting from a minor third (here $\{C - D^\sharp\}$) and going to the major third side (+4). Since $\gcd(4,12) = 3$, there are only three edges before returning to the starting minor third, so there are exactly $2 \times 3 = 6$ such chords for a given minor third, and the result is a cycle of length 6.

✻ **Third type of cycle.** Starting from a major third, take all the successive chords that are given by following the minor third side (see Figure 5.2.10): this corresponds to applying the operation $(RP)^4$ (or smply T_6^2) to a starting chord. For instance, starting the C-major chord, we get

$$\begin{aligned}(RP)^4(C) &= (RP)^3R(Cm)\\ &= (RP)^3(D^\sharp)\\ &= (RP)^2R(D^\sharp m)\\ &= (RP)^2(F^\sharp)\\ &= RPR(F^\sharp m)\\ &= RP(A)\\ &= R(Am)\\ &= C\end{aligned}$$

and we then obtain the cycle of length 8:

$$\{C, Cm\} + \{Cm, D^\sharp\} + \{D^\sharp, D^\sharp m\} + \{D^\sharp m, F^\sharp\} + \{F^\sharp, F^\sharp m\} + \{F^\sharp m, A\} + \{A, Am\} + \{Am, C\}$$

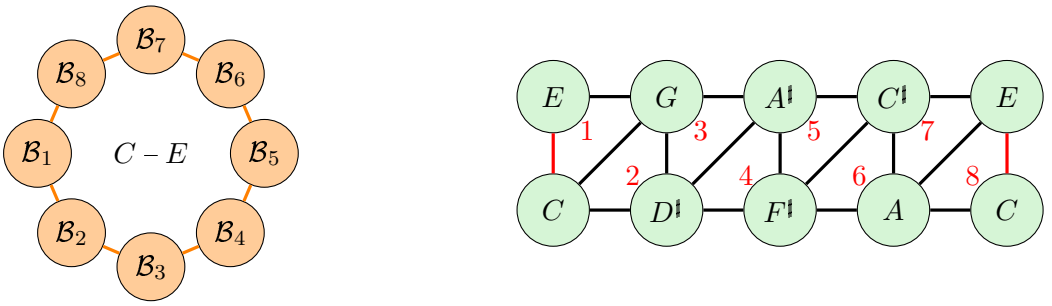


Figure 5.2.10: The homology class in degree 1 for $T[3,4,5]$, which is obtained by starting from a major (here $\{C - E\}$) and going to the the minor third (+3). Since $\gcd(3,12) = 4$, there are four edges before returning to the starting third, so there are exactly $2 \times 4 = 8$ such chords for a given minor third, and the result is a cycle of length 8.

Remark 5.2.7. We have shown how to build cycles by hand, following the sides of the minor and major thirds (+3 and +4). Notice that there is no such equivalent for the side of the fifth (+7): in fact, 7 is a generator of $\mathbb{Z}/12\mathbb{Z}$, so by following this side, we will get every note of the chromatic scale, that is every minor and major chord. On the other hand, we can create twelve cycles of the first type for each note, and twenty-four of the other two types for each chord. However, since there are several filled triangles connecting the cycles together, the dimension of H_1 is much smaller than the total number of possible cycles.

In the barcode for H_1, there are also many short bars representing some cycles of length 4: they can be considered as **noises** by looking at their lifetime in the filtration (only 1% or 2% compared to more than 10% for the cycles described below). In fact, persistent homology focuses on the bars that last during the filtration, and here the longest bars can be classified in different types of cycles, so there are relevant of some topological features about the Tonnetz, while these short bars can be considered simply as accidents.

However, there remains one bar that goes from 59% to 74%: an illustration of these two moments is given in Figure 5.2.11. This bar has many different generators considering all the filled triangles, and we can take for instance the "outline" of the graph in the illustration. In other words, this cycle could be one of the element of homology in degree 1 characterizing the torus. We also show the moment when only this cycle remains, i.e. between 71% and 73% of the filtration, which is also when the graph looks exactly like a torus.

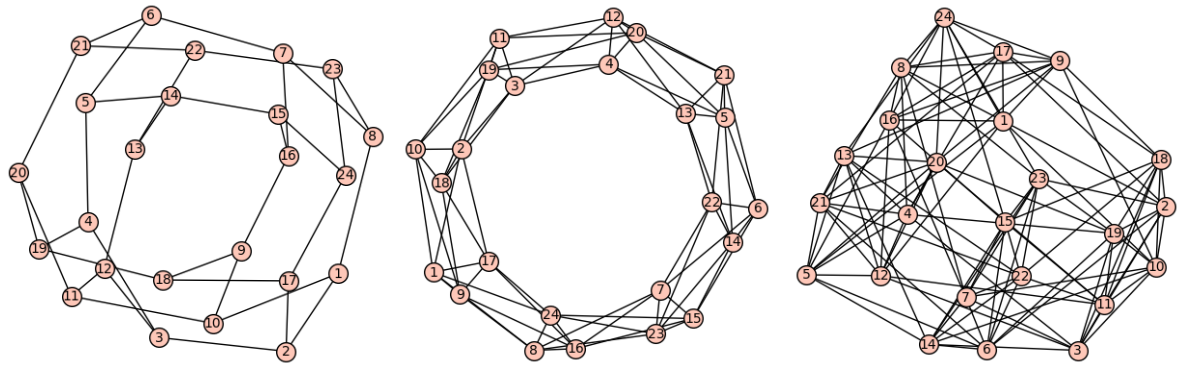


Figure 5.2.11: The filtration associated with the Tonnetz $T[3,4,5]$ with an increasing scaling parameter of 59%, 71% and 74%.

✔ **Conclusion for Euler's Tonnetz.** In the light of the above results, we can conclude that our approach is able to capture much information about the Tonnetz of minor and major chords: in degree 0 we recover the topological structure of its fundamental domain, the torus, and we also recover the Cayley group of the PLR-group. In degree 1 we obtain some interesting homology generators, which can be classified with the elements of the PLR-group and which also allow us to create one-dimensional cycles by hand. We will see in the next parts of this section that we can do the same thing with the other two-dimensional Tonnetze $T[a,b,c]$.

5.2.3. A study of the eleven other Tonnetze

Now that we have analyzed the Tonnetz $T[3,4,5]$ of major and minor chords using our DFT-distance, we can compare the work with that of the eleven other two-dimensional Tonnetze:

$$T[1,1,10],\ T[1,2,9],\ T[1,3,8],\ T[1,4,7],\ T[1,5,6],\ T[2,2,8]$$
$$T[2,3,7],\ T[2,4,6],\ T[2,5,5],\ T[3,3,6],\ T[4,4,4]$$

For this purpose, we will use exactly the same procedure as we did for $T[3,4,5]$, that means set up a bijection between the chords of $T[a,b,c]$ and the set $\{1,2,\dots,N_{a,b,c}\}$, where $N_{a,b,c}$ is the number of chords containing the studied Tonnetz. We will then create the associated score, compute barcodes and analyze the results. We will also compare them with those obtained for the Tonnetz of minor and major chords, and provide a classification of the different Tonnetze using barcodes in degree 0 and 1.

This analysis will be done using classification Table 5.2.1: indeed, we will start by studying the eight connected Tonnetze (the five tori, the two cylinders and the necklace of six tetrahedra), and then focus on the four non-connected ones (the two cylinders, the two necklaces of three tetrahedra, the three tetrahedrons and the four triangles). Also recall Remark 5.2.3 which says that there are simplicial isomorphisms between some pair of Tonnetze, given by the homotheties $z \mapsto 5z$ and $z \mapsto 7z$. More precisely, we have

$$T[1,3,8] \cong T[3,4,5], \quad T[1,2,9] \cong T[2,3,7] \text{ and } T[1,1,10] \cong T[2,5,5]$$

Since homotheties are isometries for the DFT-distance (see Theorem 2.4.14 from Chapter 2), the resulting filtations and barcodes will be exactly the same for these pairs of Tonnetze.

★ The connected Tonnetze.

Let us start with the eight connected Tonnetz:

$$T[1,1,10],\ T[1,2,9],\ T[1,3,8],\ T[1,4,7],\ T[1,5,6], T[2,3,7],\ T[2,5,5],\ T[3,4,5]$$

⋇ The five tori - ($T[1,2,9]$, $T[1,3,8]$, $T[1,4,7]$, $T[2,3,7]$, $T[3,4,5]$). We will perform exactly the same analysis as we already did for $T[3,4,5]$, and we will recover the tori given by the classification from Table 5.2.1. The analysis for $T[1,3,8]$ has already been done, since this Tonnetz is isomorphic with $T[3,4,5]$.

Let us focus on $T[1,2,9]$ and $T[2,3,7]$: the corresponding family of barcodes is shown in Figure 5.2.12, and the main steps of the filtration are illustrated in Figure 5.2.13. These different levels correspond successively to the moment when the complex is connected for the first time and when the first round of cycles appears (56%), then to the moment when the second round of cycles appears (65%) and finally to the moment when the graph changes for the first time after that second step (77%). We find back the shape of a torus at each one of these levels. Moreover, the elements of homology in degree 1 can again be classified in three types: the cycles obtained by turning around a given pitch-class (cycles of length 6) and cycles corresponding respectively to the $(+3) = (-9)$-side (length 8) and the $(+2)$-side (length 12). Note that we are now able to generate a one-dimensional cycle of length 12 by hand.

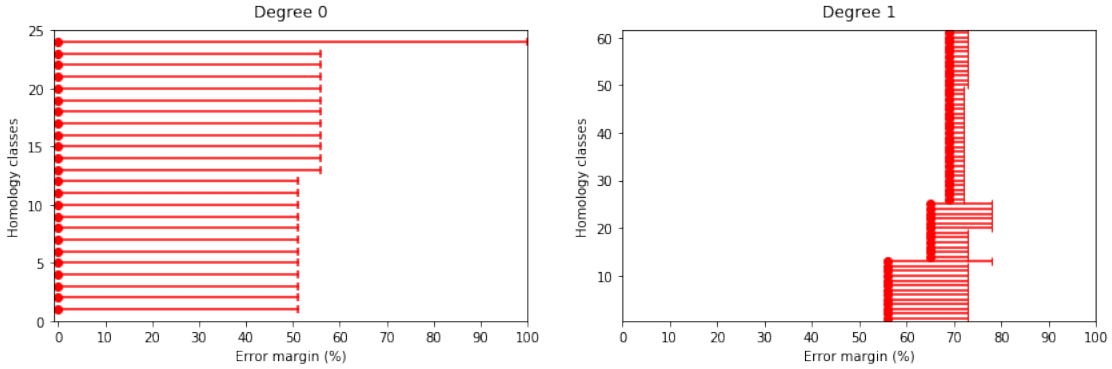


Figure 5.2.12: The associated family of barcodes with the Tonnetze $T[1,2,9]$ and $T[2,3,7]$.

Let us consider the isolated torus $T[1,4,7]$: since only one side is generated by a non-generator of $\mathbb{Z}/12\mathbb{Z}$ $(+4)$, there are only two types of one-dimensional cycles here. Actually, they are all of length 6 and are respectively given by the major third side and the action of rotating around one given pitch. We give the corresponding family of barcodes in Figure 5.2.14, and we have chosen to represent some particular moments of the filtration: here we have taken four different

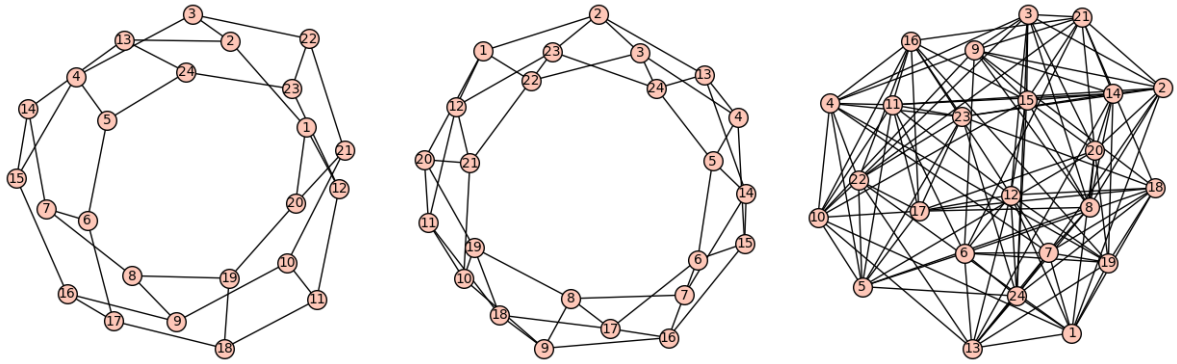


Figure 5.2.13: The filtration associated with the Tonnetze $T[1,2,9]$ and $T[2,3,7]$ with an increasing scale of 56%, 65% and 77%.

steps that are given by the first moment where the filtration is connected (59%), the moment when the second and third round of cycles appear (65% and 75%), and also the moment when the longest bar of degree 1, which lasts from from 59% to 83%, remains alone (82%). This is shown in Figure 5.2.15. The torus is clearly visible for the first two steps of the filtration.

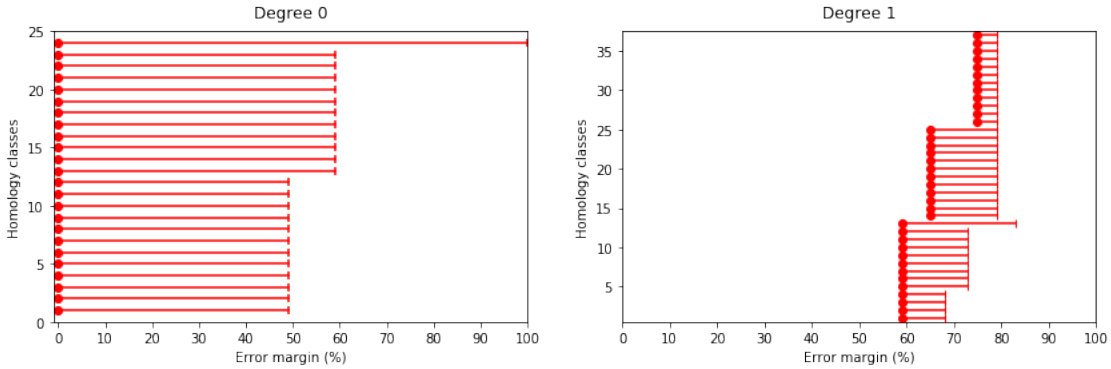


Figure 5.2.14: The associated family of barcodes with the Tonnetz $T[1,4,7]$.

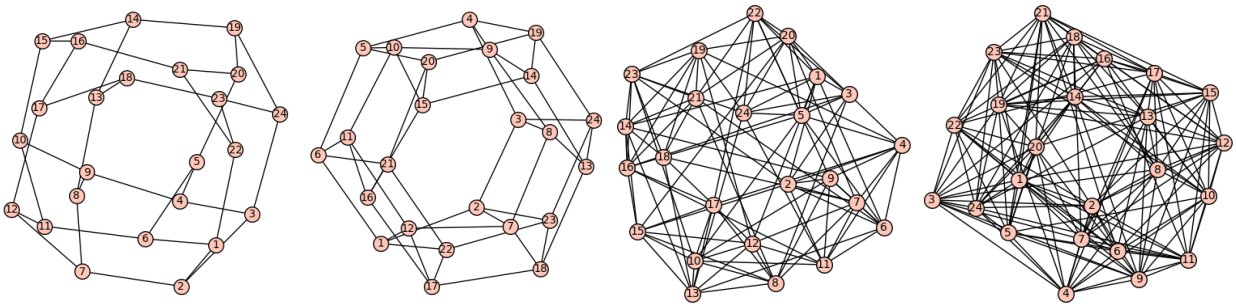


Figure 5.2.15: The filtration associated with the Tonnetz $T[1,4,7]$ with an increasing scale of 59%, 65%, 75% and 82%.

*** The two cylinders - $(T[1,1,10],\ T[2,5,5])$.** These two Tonnetze are connected and, from Table 5.2.1, they both have the topology of a cylinder. Moreover, they have twelve chords (instead of twenty-four for the tori), and are isomorphic with the homothetie $z \mapsto 5z$. The corresponding family of barcodes in degree 0 and 1 is shown in Figure 5.2.16, while some particular steps of the filtration (the first moment when it is connected and the different rounds of cycles) are illustrated in Figure 5.2.17.

For these two Tonnetze, and unlike for the tori, we are no longer able to build one-dimensional cycles by rotation: in fact, two triangles with a common edge given either by the side (+2) or the side (+10) are the same. Consequently, there are only three different chords around a given note, so we cannot form a one-dimensional cycle. On the other hand, following either the side (+2) or the side (+10) provide cycles of length 12.

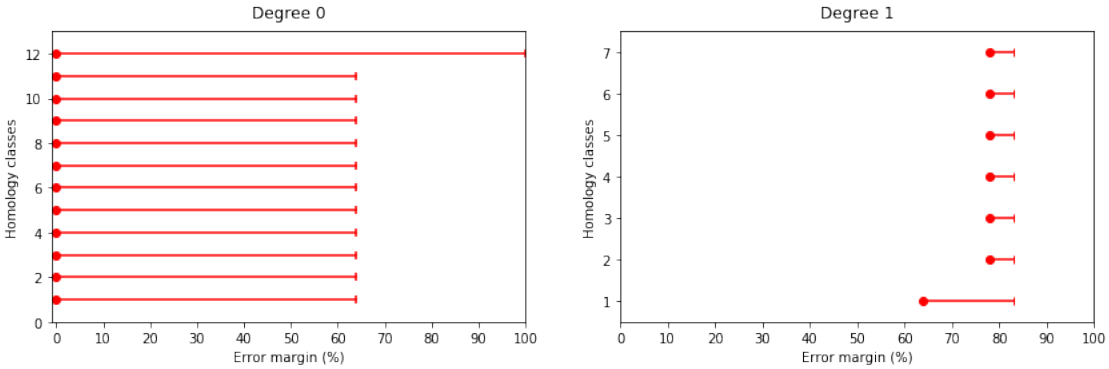


Figure 5.2.16: The associated family of barcodes with the cylinders Tonnetze $T[1,1,10]$ and $T[2,5,5]$.

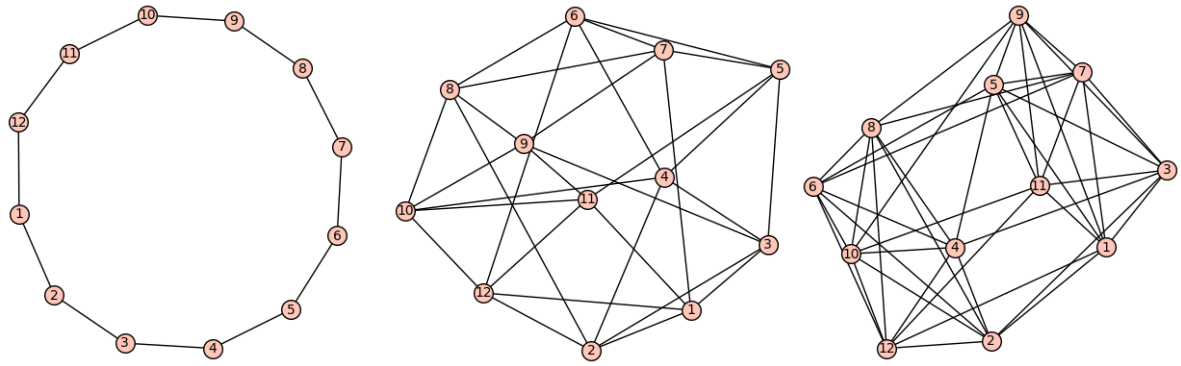


Figure 5.2.17: The filtration associated with the Tonnetze $T[1,1,10]$ and $T[2,5,5]$ with an increasing scale of 64%, 78% and 83%.

*** The necklace of six tetrahedra - $(T[1,5,6])$.** We see from Table 5.2.1 that $T[1,5,6]$ has a particular triplet of Betti numbers ($\beta_0 = 1$, $\beta_1 = 0$ and $\beta_2 = 6$), so this Tonnetz has the topology of a necklace of six tetrahedra. The corresponding family of barcodes is presented in Figure 5.2.18.

The barcode in degree 0 is different from the other connected Tonnetze: in fact, in this case we have three main filtration steps given by 46%, 53% and 58%. At 46%, half of the vertices are connected, and two vertices are linked together if they share one edge from the (+6)-side. The filtration is shown at 53% and 58% in Figure 5.2.19. In the first case, the complex is divided into two components, where each vertex has exactly three neighbors, and these components are

connected only 5% of the filtration later. With a scaling parameter of 56%, we have six groups of four vertices connected: more precisely, these are the six groups of chords that form the six tetrahedra of our necklace, so we recover the expected topological structure. We zoom in one of these tetrahedra in Figure 5.2.20, and we also put the description of each musical bar: this tetrahedron is connected to the next one by a chromatic step (+1) or a fourth (+5).

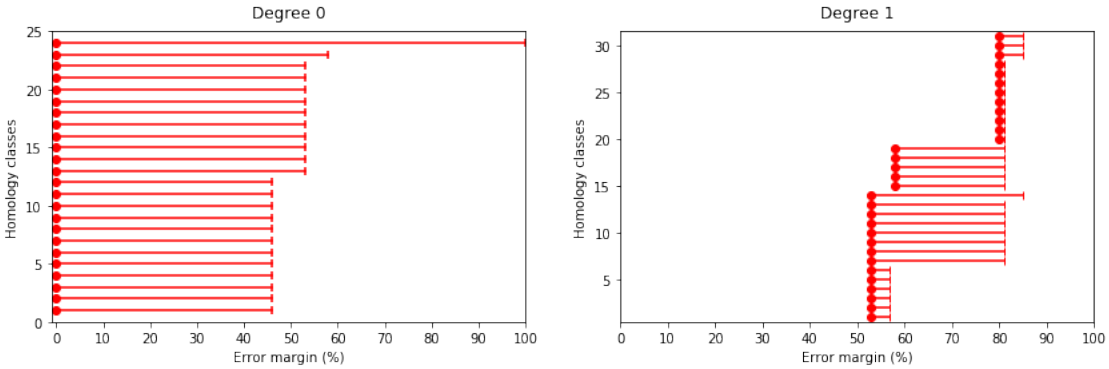


Figure 5.2.18: The associated family of barcodes with the Tonnetz $T[1,5,6]$.

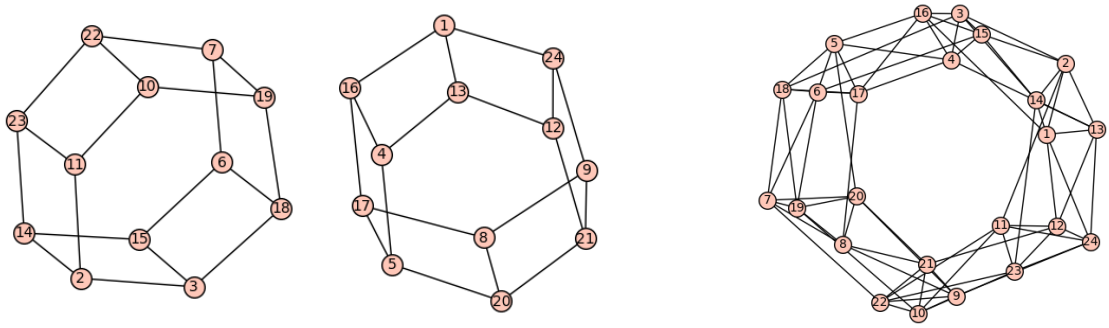
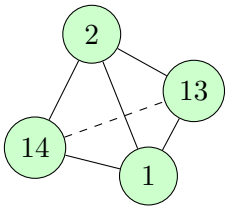


Figure 5.2.19: The filtration associated with the Tonnetz $T[1,5,6]$ at a scale of 53% and 58%. The right complex remains the same from 58% to 80% and corresponds to the necklace of six tetrahedra.



$$\begin{aligned} \mathcal{B}_1 &= \{E, F, B\} &&= \{5, 6, 11\} \\ \mathcal{B}_2 &= \{E, F, A^\sharp\} &&= \{5, 6, 10\} \\ \mathcal{B}_{13} &= \{F, A^\sharp, B\} &&= \{6, 10, 11\} \\ \mathcal{B}_{14} &= \{A^\sharp, B, E\} &&= \{10, 11, 5\} \end{aligned}$$

Figure 5.2.20: A zoom to one of the six tetrahedra of the Tonnetz $T[1,5,6]$.

Concerning degree 1, $T[1,5,6]$ is the only connected Tonnetz where we do not have a natural way to produce one-dimensional cycles. In fact, two sides of this Tonnetz are generators of $\mathbb{Z}/12\mathbb{Z}$ and one has period two, so the triangles are all very close in terms of distances. However, we see that the barcode is far from being empty and that there are indeed different types of cycles that last during the filtration: in fact, some of them last only 5% (from 53% to 58%) and correspond to the twelve outlines (cycles of length 4) that we can clearly see on the graph of Figure 5.2.19. At 58%, half of them disappear because they are connected by boundaries, so we still have one-dimensional cycles of length 4. On the other hand, we also get new generators

which are given by the fourth side, but in a special way which is specific to this Tonnetz and shown in Figure 5.2.21. In fact, we see that this cycle has length 10 and is glued together to one of the above cycles of length 4. Also notice that this type of one-dimensional cycle appears only twice in this Tonnetz: in fact, it is generated by twelve chords and the Tonnetz $T[1,5,6]$ contains twenty-four ones. These two cycles are linked together by boundaries and they can be given by the outline of the necklace. Nevertheless, we obtain with $T[1,5,6]$ a new family of one-dimensional cycles of length 10.

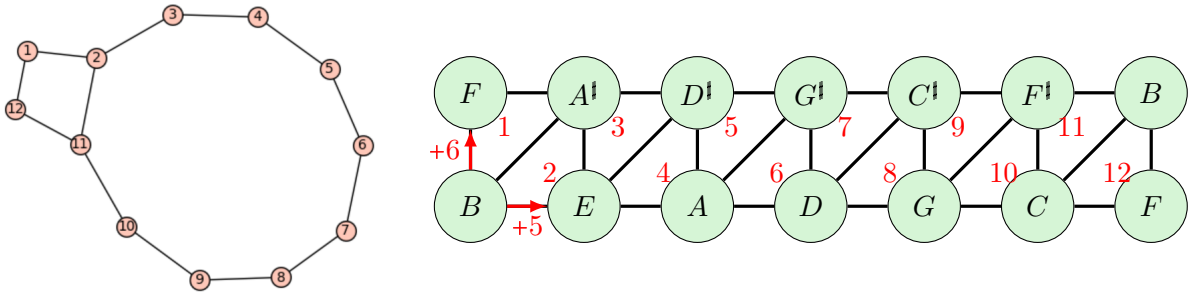


Figure 5.2.21: The one-dimensional cycle of length 10 which is given by the Tonnetz $T[1,5,6]$.

★ The non-connected Tonnetze.

We will now complete this analysis by looking at the four non-connected Tonnetze:

$$T[2,2,8],\ T[2,4,6],\ T[3,3,6],\ T[4,4,4].$$

The first observation we can make is that some of these Tonnetze contain very few different chords (four, twelve and twenty-four) so we will not get one-dimensional cycles as easily as we did for the connected ones. However, we have seen in the previous analysis that the barcodes in degree 0 provide a good illustration of the topological structure of the Tonnetz, so we expect to find here the different connected components that actually compose these different Tonnetze.

⋇ The two cylinders - $(T[2,2,8])$. This Tonnetz contains only twelve chords and has two similar generators, given by the major second side $(+2)$. Consequently, two triangles that share a $(+8)$-side are the same. Thus, as for the two cylinders $T[1,1,10]$ and $T[2,5,5]$, we will not have one-dimensional cycles by turning around a given note.

Let us focus on the barcode in degree 0, as shown in Figure 5.2.22. We can immediately see that there are two general steps in this graph: the first one, where only two components remain (59%), and the second, which comes much more later, with a scaling parameter of 92%. These two moments are shown in Figure 5.2.23. Moreover, the two components at 59% are exactly given by the two cylinders and the filtration remains divided for more than 30% of the filtration, which we never had with the connected Tonnetze. Therefore, the persistent interpretation of the degree 0 tells us that this set of 3-chords is divided into two different parts. Furthermore, these two parts are also the two one-dimensional cycles of the barcode in degree 1, which are of length 6. This analysis shows us that we have recovered the non-connected characteristic from this Tonnetz and also a new way of creating one-dimensional cycles of length 6.

⋇ The two necklaces of three tetrahedra - $(T[2,4,6])$. This Tonnetz contains twenty-four chords, and the associated family of barcodes is shown in Figure 5.2.24. With this Tonnetz, we still get the idea that the barcode in degree 0 reveals two general steps of the filtration: the one at 53%, when the twenty-four chords are divided into two parts and thus provide the two different necklaces, and the one at 82%, when these two components are grouped. These two moments are shown in Figure 5.2.25.

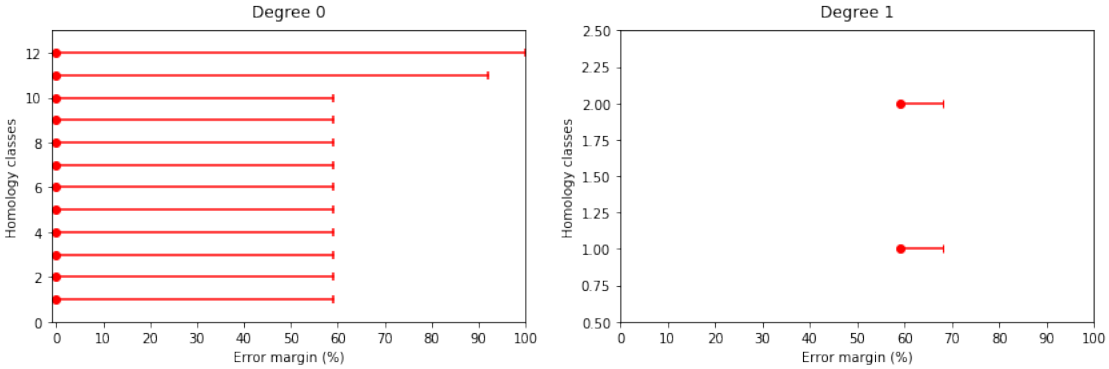


Figure 5.2.22: The associated family of barcodes with the Tonnetz $T[2,2,8]$.

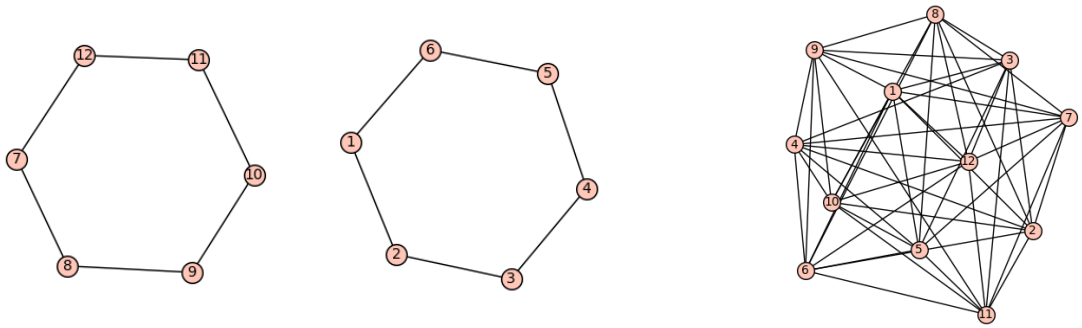


Figure 5.2.23: The filtration associated with the Tonnetz $T[2,2,8]$ with a scaling parameter of 59% (left) and 92% (right). The left complex remains the same from 59% to 68%, which is also the length of the two one-dimensional cycles.

The analysis of the one-dimensional cycles is roughly similar to the one for the necklace $T[1,5,6]$: indeed, there are cycles of length 4 that are given by the twelve outlines that we can clearly see on the graph at a scale of 53%. At 58%, half of these cycles disappear because they become boundaries. By using the major third side, we also get cycles of length 6. In degree 1, we can see that two bars rise when the complex is connected, with a scaling parameter of 82%: as it already was the case before, these two cycles are given by the outline of the graph.

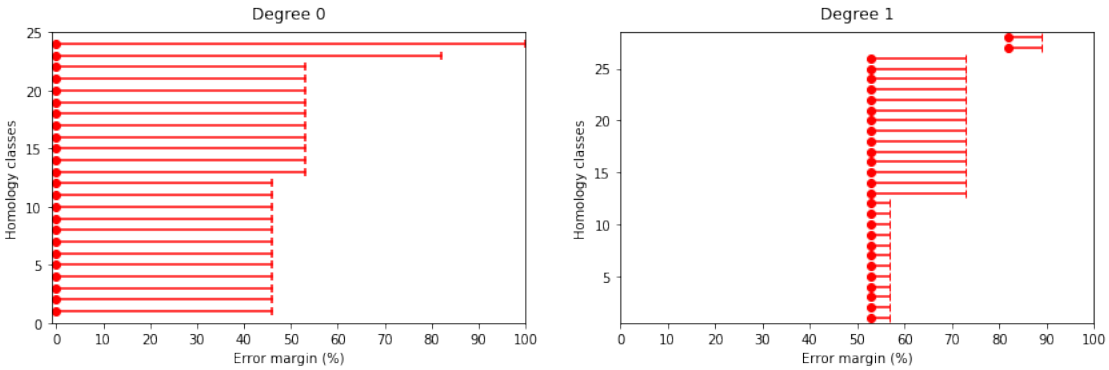


Figure 5.2.24: The associated family of barcodes with the Tonnetz $T[2,4,6]$.

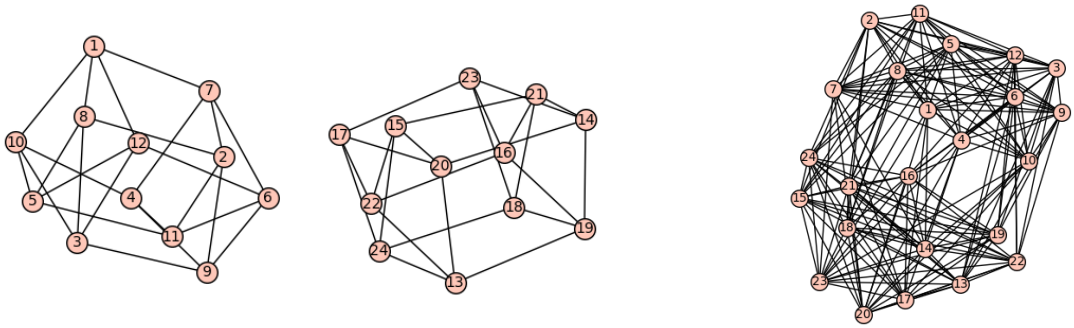


Figure 5.2.25: The filtration associated with the Tonnetz $T[2,4,6]$ from the moment when the two necklaces appear (53%) and the one when it is connected (82%).

*** The three tetrahedra - ($T[3,3,6]$).** This Tonnetz contains only twelve chords and is generated by two similar sides, the minor third (+3). The associated family of barcodes in degree 0 and 1 is given in Figure 5.2.26.

For this Tonnetz, whose fundamental domain is divided into three parts each containing four chords, we find two main steps of filtration in degree 0: at 65%, when the twelve vertices are divided into three connected graphs (the three tetrahedra), and at 93%, when these graphs are glued together. These two moments are illustrated in Figure 5.2.27. Notice that this graph is in three parts for almost 30% of the filtration. It is also at the end of this moment that the one-dimensional cycles appear, with the connection of the tetrahedra forming cycles of length 4. However, the barcode in degree 0 again confirms that this Tonnetz is in three different parts.

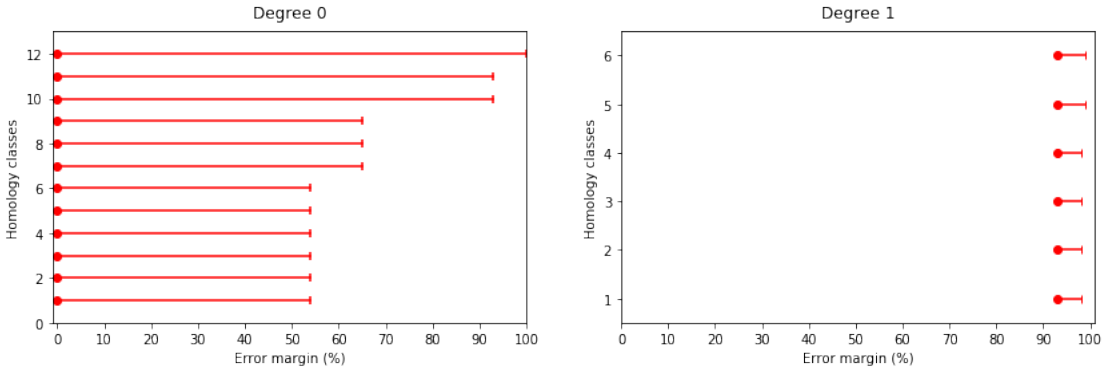


Figure 5.2.26: The associated family of barcodes with the Tonnetz $T[3,3,6]$.

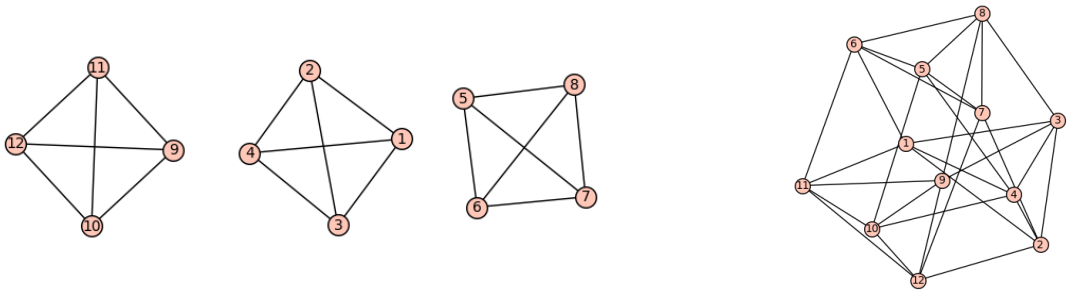


Figure 5.2.27: The filtration associated with the Tonnetz $T[3,3,6]$ from the moment when the three tetrahedra appear (65%) and the one when it is connected (93%).

❋ **The four triangles - ($T[4,4,4]$).** This Tonnetz is the least interesting, since it is formed with the same side (+4), so there are only four chords in it. These four chords are all obtained by a simple 1-transposition, as shown below:

$$\begin{array}{rcccl} \mathcal{B}_1 & = & \{C, E, G\sharp\} & = & \{0,4,8\} \\ \mathcal{B}_2 & = & \{C\sharp, F, A\} & = & \{1,5,9\} \\ \mathcal{B}_3 & = & \{D, F\sharp, A\sharp\} & = & \{2,6,10\} \\ \mathcal{B}_4 & = & \{D\sharp, G, B\} & = & \{3,7,11\} \end{array}$$

However, if there are no one-dimensional cycles given the number of musical bars, the corresponding family of barcodes, shown in Figure 5.2.28, shows us that the Tonnetz $T[4,4,4]$ is separated for most of the filtration time, more precisely until 82%. At this moment, two chords are connected until 100% of the filtration, when we finally get the complete graph. These two moments are shown in Figure 5.2.29, and it confirms that this Tonnetz is well composed of distinct components.

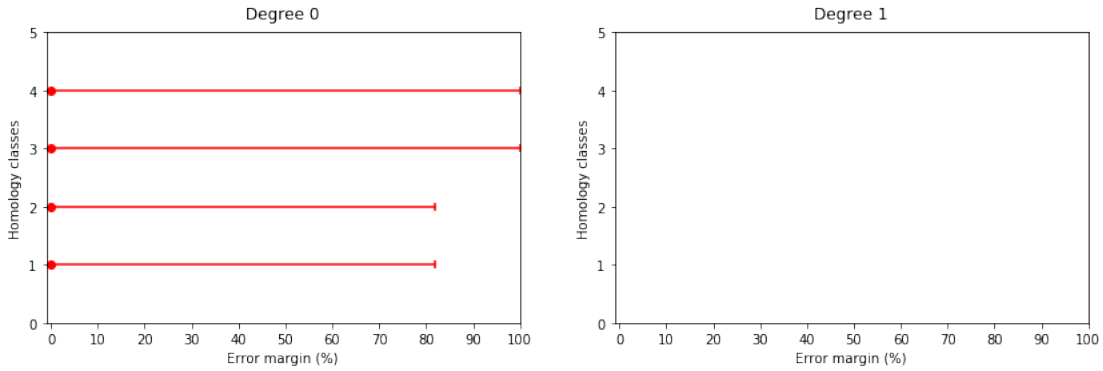


Figure 5.2.28: The associated family of barcodes with the Tonnetz $T[4,4,4]$.

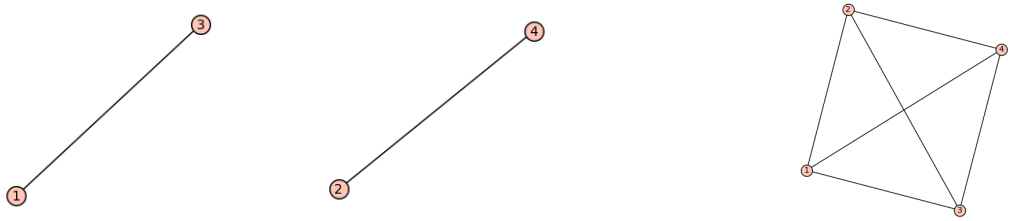


Figure 5.2.29: The filtration associated with the Tonnetz $T[4,4,4]$ from the moment when the first chords are connected (82%) and the one when it is fully connected, which is simply the end of the filtration (100%).

5.2.4. A classification of the Tonnetze using the DFT

We conclude with a summary of what we have done in this section: we decided to apply our DFT-distance on different sets of 3-chords, which we selected using the different two-dimensional Tonnetze. This analysis shows that barcodes in degree 0 allow us to recover the topological structure of each Tonnetz. Note that we have arbitrarily chosen a moment of filtration (most of the time, this was the first moment when the filtration was connected) and we will refine this choice in the next chapter (see Section 6.4). In particular, Table 6.4.4 proposes a "graph-type" for each Tonnetz and scale studied. For each graph-type, we will recover the different characteristic shapes we found in this first analysis.

In addition, barcodes of degree 1 allow us to manually generate one-dimensional cycles. Each Tonnetz provides specific types of generator of H_1, and it is interesting to compare them: for some Tonnetz we have cycles of length 4, while for others it can go up to 12. Table 5.2.3 classifies the different types of cycles we can find in these two-dimensional Tonnetze.

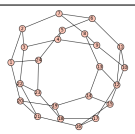
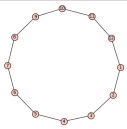
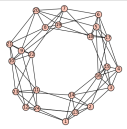
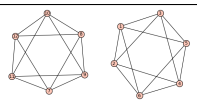
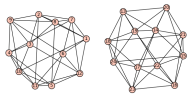
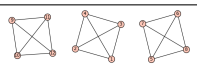
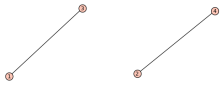
Tonnetz	Topology (degree 0)	Lengths of H_1-cycles				
		4	6	8	10	12
$T[1,2,9]$, $T[2,3,7]$			×	×		×
$T[1,3,8]$, $T[3,4,5]$			×	×		
$T[1,4,7]$			×			
$T[1,1,10]$, $T[2,5,5]$						×
$T[1,5,6]$					×	
$T[2,2,8]$			×			
$T[2,4,6]$		×	×			
$T[3,3,6]$		×				
$T[4,4,4]$						

Table 5.2.3: A classification of the twelve two-dimensional Tonnetz $T[a,b,-(a+b)]$ by their topological structure and their respective types of cycles in H_1 found using the DFT.

In conclusion, this first analysis confirms our belief that the distance constructed by means of the DFT together with persistent homology seems to be a reasonable tool for understanding known musical structures.

Chapter 6.

Harmonization of Pop songs

6.1. General idea

The main idea of this chapter is to look at music pieces built on a certain number N of chords: indeed, if the general purpose of using persistent homology in music analysis is to apply it on any possible piece of music, we might want to have a deep understanding of what happens on very basic examples. In that order, we have chosen to apply our TDA approach together with the two-dimensional DFT to the most possible **reduced songs**.

The process is the following: let us consider a music piece $\mathfrak{P}$, which is assumed to be extracted from the large database of Pop music. In other words, $\mathfrak{P}$ represents a **song**, with different type of musical instruments: voice, piano, guitar, bass, drums, etc. Here we are mostly interested in the accompaniment, and more precisely the chord chart, so we are going to create the associated score $\mathcal{S}_{\mathfrak{P}}$ made of **two tracks**:

- First track: the **melody** played as in the original score, living in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ for a well-chosen time-unit t and a pitch-unit p.
- Second track: the **accompaniment** with chords in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$ and the minimal time-unit is always a quarter note ♩, so the accompaniment is as minimal as possible.

We call this process the **harmonization of a song**. We will then have several level of analysis, depending on what we are looking at between the chord chart, the accompaniment or the global harmonized song. We are now going to illustrate this process with a basic example.

Example 6.1.1. Let us consider the song *Forever Young* from the Pop group Alphaville (1984). This song is based on the five following chords:

$$C - G - Am - F - Dm.$$

Now, if we only keep the melody and the accompaniment with quarter notes, we get the score that is presented in Figure 6.1.1, which illustrates the first theme of the song.

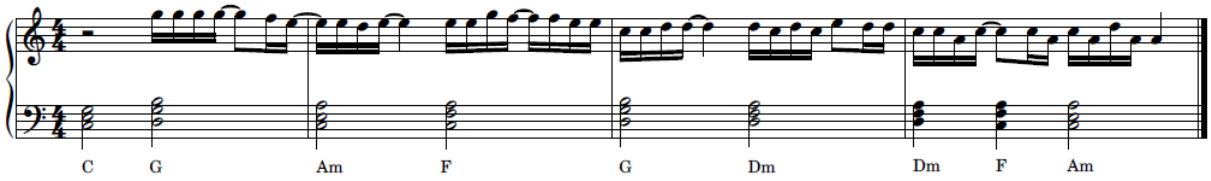


Figure 6.1.1: An excerpt of the harmonization of the pop song *Forever Young* from the group Alphaville (1984). This song is based on the five chords $C - G - Am - F - Dm$.

The idea of reducing a music piece to the most simple structure (melody + accompaniment) comes naturally in the idea of understanding our new process of musical analysis: in fact, it allows us to have a deeper overview of how it works in reduced case before extending it to more general contexts.

Moreover, in our case this approach has a further justification especially after the works that was done on the different Tonnetze in Section 5.2: in fact, we saw that the DFT, together with persistent homology, has a worthwhile musical interpretation when it is applied on chords, so we expect it to work particularly well for this construction.

In that purpose, we propose a protocol made of different and progressive levels of analysis: each level will correspond to a way of extracting a filtered complex and a family of barcodes from the original score based on the Fourier point cloud method (see Definition 4.2.3). Then, the idea is to compare barcodes and complexes between the different level of analysis, and finally between several songs. For each level, we will illustrate our analysis with the harmonized song *Forever Young* from Example 6.1.1.

⁕ Three levels of analysis for harmonized Pop songs.

- **Level 0**: Only look at the chord chart, that means that a chord of the song is a musical bar, which contains only this chord (played with whole notes, as we did for the Tonnetze). Then, each bar is living in $\mathbb{Z}/1\mathbb{Z}\times\mathbb{Z}/12\mathbb{Z}$ and we thus get complexes with N vertices, where N is the number of chords into the chord chart. For instance, in the case of *Forever Young*, the result is the score from Figure 6.1.2, and it leads to a filtered complex with only five vertices. Thus, we build the associated filtration and compute persistent homology, so it allows us to compare the different chord charts that are commonly used in Pop music.

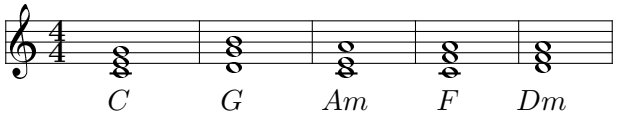


Figure 6.1.2: The chord chart of *Forever Young* from Alphaville (1984). Since each chord from $\{C, G, Am, F, Dm\}$ represents a musical bar, the associated filtered complex has five vertices.

- **Level 1**: Study the accompaniment in itself, that means taking into account the rhythms and also the chord's order. Since we reduced the score at much as possible, we consider that for every song the time-unit for the accompaniment is measured with a quarter note. A chord can thus be placed in n different places in the bar, where n is the number of beats for the song (for instance 4 if the meter is 4/4), and we are looking at the different configurations we might have. For instance, some songs could have all their chords in the first time of the bar so the filtration will be the same at level 0 and level 1, while some others will use various permutations of chords and have a *richer* and more *complex* accompaniment. Figure 6.1.3 shows an excerpt of the accompaniment of *Forever Young*. If this song contains only five different chords in its chord chart, at level 1 we have nine distinct musical bars, given by:

$$\begin{aligned}
\mathcal{B}_1 &= \{(0,C),(16,G)\} \\
\mathcal{B}_2 &= \{(0,Am),(16,F)\} \\
\mathcal{B}_3 &= \{(0,G),(16,Dm)\} \\
\mathcal{B}_4 &= \{(0,F),(16,Am)\} \\
\mathcal{B}_5 &= \{(0,Dm),(8,F),(16,Am)\}
\end{aligned}
\qquad
\begin{aligned}
\mathcal{B}_6 &= \{(0,F),(16,G)\} \\
\mathcal{B}_7 &= \{(0,Am),(16,C),(24,G)\} \\
\mathcal{B}_8 &= \{(0,F),(16,C)\} \\
\mathcal{B}_9 &= \{(0,C)\}
\end{aligned}$$

With this description, it is implied that each point is actually a 3-chord, so for instance $\mathcal{B}_1$ does not contains two but six notes. If the beat unit is given by a quarter note for the

accompaniment, in that song the shortest note for the melody is a sixteenth note so the time-unit will be given by $t = 32$. The half note value is thus equal to 16 and the quarter ones to 8, which is why the accompaniment is encoded as above.

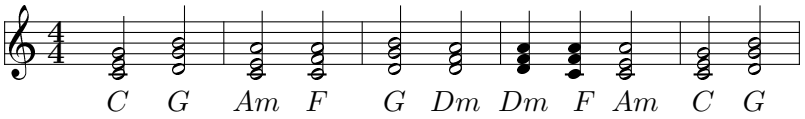


Figure 6.1.3: An excerpt of the accompaniment of *Forever Young* from Alphaville (1984). Here we have five different musical bars, where the five chords of $\{C, G, Am, F, Dm\}$ are distributed according to the four possible positions.

In this example, we started with a song with five different chords in its chord chart and still obtain an accompaniment which contains nine different musical bars: in terms of filtration, we go from one filtered complex with five vertices to one that has nine. Moreover, this complex has vertices of the form

$$t_1.Ch_1 \ - \ \ldots \ - \ t_n.Ch_n$$

with n being the number of chords Ch_i in the bar and t_i the corresponding onset. For instance, in this example the first bar $\mathcal{B}_1$ will be represented by the vertex

$$0.C \ - \ 16.F.$$

This level provides a natural axis of study: comparing these two levels of analysis in order to see what the rhythms and chords ordering bring to the song, and how our DFT-distance analyzes it.

- **Level 2**: This level studies the global reduced song in itself, that means the melody in the first voice and the accompaniment in the second. It allows us to see what the melody brings to the song: in fact, some songs could have a very basic accompaniment and still have a *complex* melody, in a sense that needs to be precise. In the case of *Forever Young* from Alphaville at that level of analysis, we go back to the score from Figure 6.1.1.

Remark 6.1.2. If we are particularly interested in the chord chart and the accompaniment here, we might also want to study the melody of a given song as a score itself, and thus conversely see what the accompaniment brings to this melody, or even do some comparison between several melodies (with and without accompaniment). This could be a trail for future possible work.

Also notice that, in order to reduce the songs as much as possible, we will restrict our work to songs built only on **major and minor chords**, that is, chords from the Tonnetz $T[3, 4, 5]$. This will allow us to look at the chord chart into this Tonnetz. Again, an idea for extending this work might be to work with more general types of chords.

For each level of analysis, we will look at several *key moments* of the filtration: in degree 0, we arbitrarily choose in a first place to look at **the scale for which the complex is connected**. In fact, this moment can be considered as a well-chosen time of analysis, especially after the classification we made for the Tonnetze, where the first connected time allowed us to recognize the topological structure (see Section 5.2). We will then refine this moment in a second place.

In the different description of the analysis levels, we understood that an important part of this section consists on comparing the different levels together. In particular, we are interested in the comparison between levels 0 and 1: if we take two songs that are built on the same chord chart, for instance the famous

$$C \ - \ G \ - \ Am \ - \ F$$

or any of its transposition, we will have exactly the same filtration and family of barcodes in level 0, but we also might get a different filtration in level 1. More precisely, the filtration for a song in level 1 might seems more complex than the one for another song. This kind of comparison will lead to the definition of **complexity of the accompaniment**, that allows us to classify several songs together (see Paragraph 6.4.3).

Furthermore, we will propose a way to compare level 1 with level 2 by looking at both filtrations and adding colors to the second graph according to its accompaniment, that means gathering the musical bars that share the same accompaniment. In other words, we will study the **bundles**, and this will allows us to analyze the behavior of the melody according to the chords, and how the DFT gathers the musical bars.

Remark 6.1.3. By studying songs based on N chords for small values of N (2, 3, 4, 5 and 6), we will see that it often yields to some recognizable filtrations, that we may want to study manually by visualizing them with graphs. Thus, as we already discussed in Remark 4.1.3 when we defined the Vietoris-Rips filtration, it will seem natural to look at the homology of the graphs instead of the complexes, so we will naturally construct a method to associate a **graph-type** to any pop song built by this harmonization process, and we will also extend it in more general cases. In particular, we will apply it to the previous work with scales and the Tonnetz, and recover the different results we have found in those sections.

For all the following analyzes, we say that a song based on N chords is a **N-chords song**, and we have created corresponding MIDI files that are available on the online database:

`https://math-musique.pages.math.unistra.fr/midi.html`.

6.2. Borderline cases: two and three chords

In this section, we start by analyzing the most simple case, that means Pop songs that are built on two or three chords. This case shows the relevance of this analysis and especially the comparison part between level 0 and level 1: in fact, the filtration associated with the chord chart will eventually go to an edge or a triangle, and there is no other choice, while level 1 might be even more complex than that.

6.2.1. 2-chords songs

We start by taking three songs from three different groups and which are built on only two chords: *Something in the Way*, *Born in the USA* and *Eleonar Rigby*. The songs are presented on Table 6.2.1.

Song	Artist	Year	Chord chart
Something in the way	Nirvana	1991	$F\sharp m$ – D
Born in the USA	Bruce Springsteen	1984	E – B
Eleanor Rigby	The Beatles	1966	C – Em

Table 6.2.1: The list of the three studied 2-chords songs.

We can immediately notice that the choice of the two chords on which the songs are built is not systematic: in fact, the songs by The Beatles and Nirvana are built on a major and a minor chord, and more precisely on chords that are linked by a leading-tone transformation (see Section 5.2.2, Definition 5.2.4). On the other hand, *Born in the USA* is built on two major chords, i.e. chords that are separated at least from two transformations on the Tonnetz $T[3,4,5]$: more precisely, we use a transformation LP to go from E to B.

Notice that, for these three songs, the level 0 of analysis will be the same, that means that the filtration will be in two components (one vertex for each chord), and it will be non-connected until the end of the filtration (100%). The most interesting aspect here is the difference between level 0 and level 1.

*** Something in the Way.** This song from the group Nirvana is built on the two chords $F^\sharp m$ and D that are distant from a leading-tone transformation on the Tonnetz $T[3,4,5]$. Thus, the complex at level 0 (chord chart) has only two vertices that are two components until the end of the filtration. At level 1, we are in the most simple case where the accompaniment is only given by whole notes: in consequence, the filtration is the same for both levels, and the family of barcodes is given in Figure 6.2.1. The only difference goes with the labels of vertices: at level 0, the complex has two vertices given by $\{F^\sharp m\}$ and $\{D\}$, while at level 1 the vertices are of the form $\{0.F^\sharp m\}$ and $\{0.D\}$.

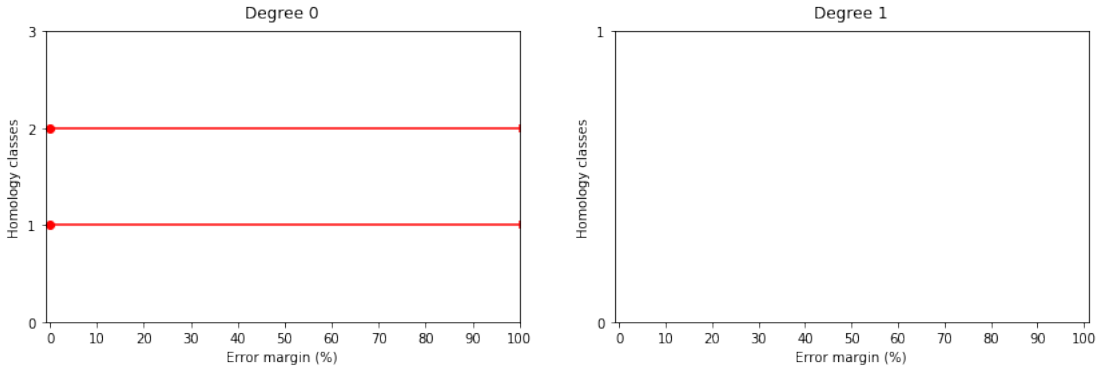


Figure 6.2.1: The associated family of barcodes for the harmonization of *Something in the Way* for levels 0 and 1 of analysis (the filtrations are the same in both cases).

For level 2 of analysis, we add the melody and get nineteen distinct musical bars so the filtration has nineteen vertices, and each of them has a color of type $\{F^\sharp m\}$ or $\{D\}$. The associated family of barcodes is given in Figure 6.2.2. Moreover, we chose to represent the filtration for each level in Figure 6.2.3 at a scale of 61%: in fact, because here level 0 and level 1 are the same and the associated filtrations are connected at 100%, we only represented the moment where the complex in level 2 is connected, that means at 61%. In that colored complex, we see that there is a clear separation between the musical bars according to their accompaniment, which seems to confirm the benefits of taking the DFT as a metric.

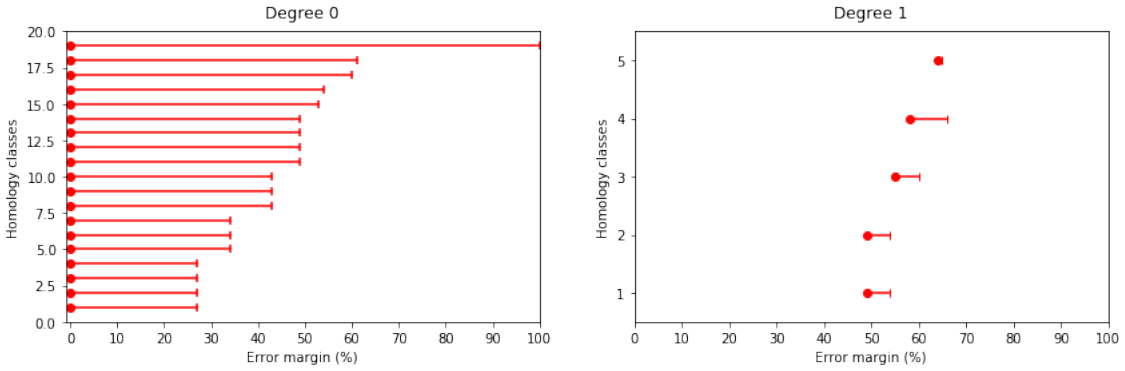


Figure 6.2.2: The associated family of barcodes for the harmonization of *Something in the Way* at level 2 of analysis.

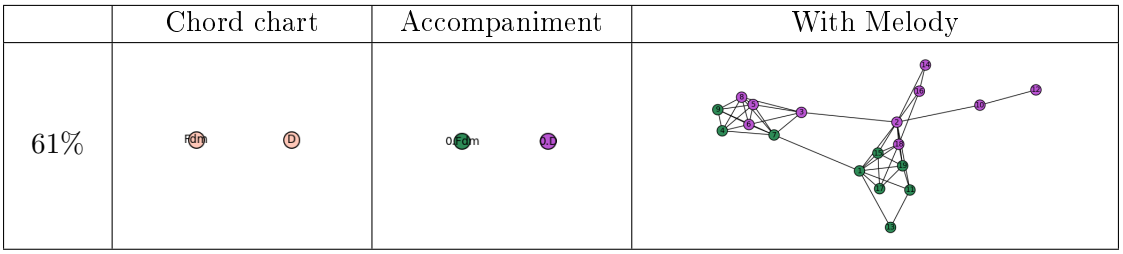


Figure 6.2.3: Comparison of the three levels of analysis for *Something in the Way* at a scale of 61%. We can observe at level 2 the clear separation between the two types of musical bars according to their accompaniment.

❋ **Born in the USA.** The analysis of this song by Bruce Springsteen is roughly similar to the previous one, at least for the first levels: in fact, the levels 0 and 1 are the same (each chord falls at the first time of the musical bar) and the associated family of barcodes is thus exactly the same as for Nirvana (Figure 6.2.1).

However, we have chosen to study this song because of the difference that brings the melody compare to *Something in the Way*: in fact, if levels 0 and 1 are the same for both songs, the first is a slow ballade and the melody is repetitive (which results in a small numbers of musical bars), so the complex seems more "simple". On the contrary, if *Born in the USA* counts only two chords in its chord chart, the harmonized song counts forty-eight distinct musical bars and seems more complicated according to the resulting complex that we can see in Figure 6.2.5. This difference of "complexity" can also be observed in both barcodes in degree 1: in fact, the number of generator of homology are not the same and there is much more in the second song. The different bars are also longest for *Born in the USA*. We thus see with this example that two songs with the same filtrations at levels 0 and 1 can be differentiated by the third level of analysis. Moreover, we still observe for this song the separation between the two colored kinds of musical bars at level 2.

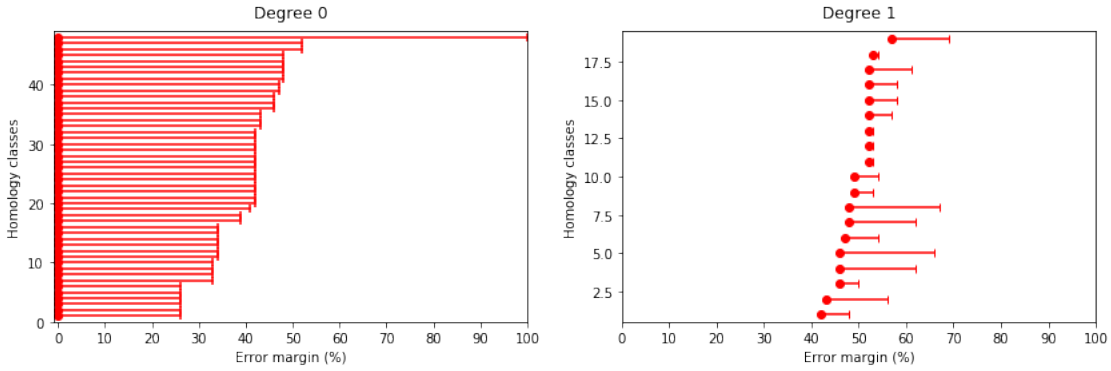


Figure 6.2.4: The associated family of barcodes for the harmonization of *Born in the USA* at level 2 of analysis.

❋ **Eleanor Rigby.** This last song we chose to study is representative of what is possible to do with only two chords: in fact, if the family of barcodes are the same as for *Something in the Way* and *Born in the USA* at level 0 (Figure 6.2.1), one can observe the real difference at level 1. The associated family of barcodes is presented in Figure 6.2.6. Instead of having only two distinct musical bars, with that song we have now nine distinct ones, and the filtration is thus much more complicated than the ones with the previous songs at the same level of analysis. There are also two long bars in degree 1, with one that lasts almost 20% of the filtration, which is quite a

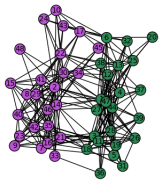
	Chord chart	Accompaniment	With Melody
62%			

Figure 6.2.5: Comparison of the three levels of analysis for *Born in the USA* at a scale of 62%. We can observe at level 2 the clear separation between the two types of musical bars according to their accompaniment.

while. On the contrary, the melody is much more repetitive for this song than for *Born in the USA*, and the complex in level 2 counts only twenty-five distinct musical bars. The difference of complexity can be observed by comparing both barcodes in degree 1, where we can see that there are much less bars for *Eleanor Rigby*.

Let us briefly focus on the filtration for *Eleonar Rigby* at level 1. As for the two previous 2 chords songs, we compare the three levels of analysis in Figure 6.2.8: the difference here is that we represent the moments when level 1 and level 2 are connected (respectively 62% and 60%). At a scale of 60%, the complex at level 1 has four components and each component is a "type of accompaniment", according to the position of the chord Ch_i and Ch_j:

$$0.Ch_1 \ - \ 0.Ch_2 \quad ; \quad 0.Ch_i \ - \ 4.Ch_j \quad ; \quad 0.Ch_i \ - \ 8.Ch_j \quad ; \quad 0.Ch_i \ - \ 12.Ch_j$$

At 62%, all these components are connected. At level 2, the complex has nine different colors and eleven vertices out of twenty-five have color associated with the musical bar $\{0.Em\}$. It turns out that the tonality of *Eleanor Rigby* is Em, which means that this new information can now be found out automatically through this analysis.

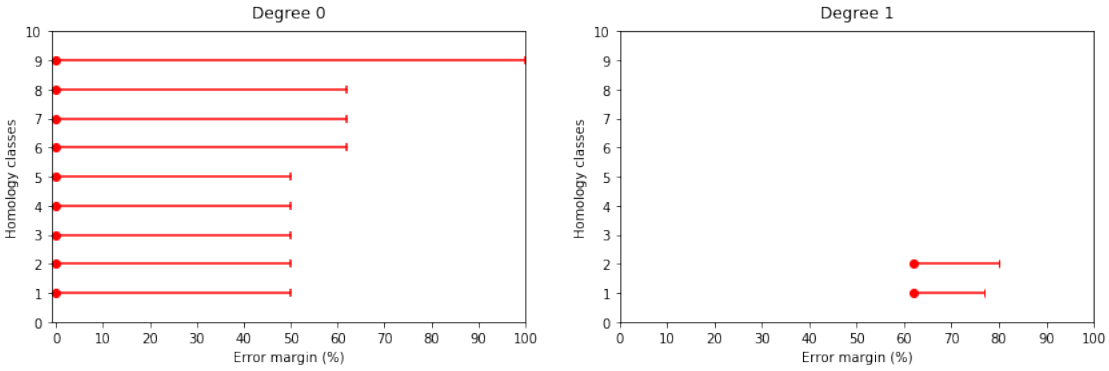


Figure 6.2.6: The associated family of barcodes for the harmonization of *Eleanor Rigby* at level 1 of analysis.

✔ **Conclusion for 2-chords songs.** We saw here three different kinds of songs that are all built on only two chords and each level of analysis revels different musical features. For *Something in the Way* and *Born in the USA*, the first two levels of analysis are the same but the melody brings more complexity for the second. In both cases, the musical bars are gathered by accompaniment. For *Eleanor Rigby*, if the general filtration at level 2 is less complex than *Born in the USA*, the one at level 1 reveals another kind of complexity that is also visible by looking at the barcodes in degree 1. The general comparison of these three songs is illustrated in Figure 6.2.9.

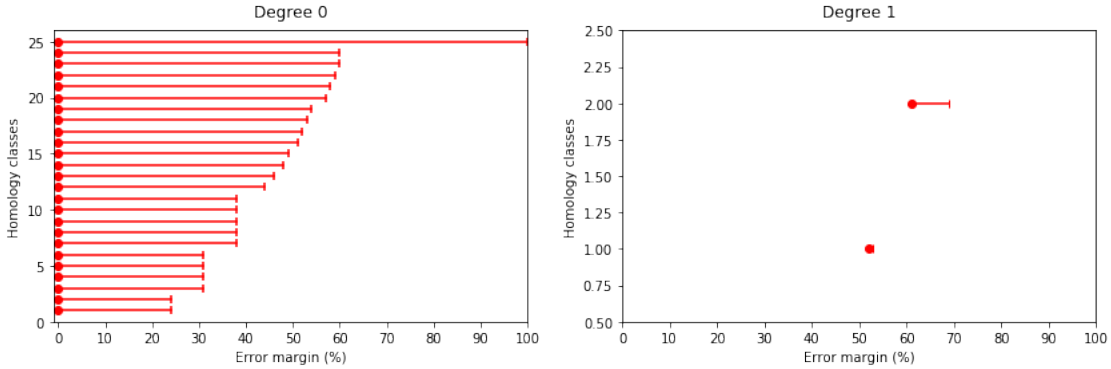


Figure 6.2.7: The associated family of barcodes for the harmonization of *Eleanor Rigby* at level 2 of analysis.

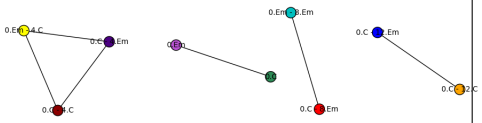
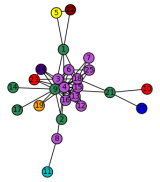
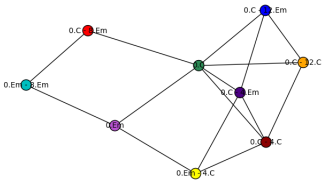
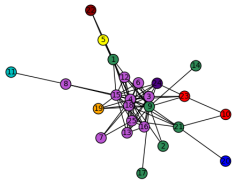
	Chord chart	Accompaniment	With Melody
60%			
62%			

Figure 6.2.8: Comparison of the three levels of analysis for *Eleanor Rigby* with a scale respectively of 60% and 62%. We can observe the complexity of the filtration at level 1, and the majority of vertices have color $\{0.Em\}$ at level 2, which indicates that the song tonality is E_m.

6.2.2. 3-CHORDS SONGS

We will now consider three songs from three different groups that are built this time on three chords: *What's up*, *I have a dream* and *I love Rock N'Roll*. The songs are presented on Table 6.2.2.

Song	Artist	Year	Chord chart
What's up	4 Non Blonds	1993	$A - Bm - D$
I have a dream	ABBA	1979	$B\flat - F - E\flat$
I love Rock N'Roll	Joan Lett	1981	$E - A - B$

Table 6.2.2: The list of the three studied 3-chords songs.

As for the 2-chords songs, the choice of the chord chart is not systematic: for instance, the first song have two major chords and one minor, while the two other songs have only major chords. However and unlike the previous paragraph, this time the chord chart will influence the filtration at level 0 of analysis. In fact, since now we have three chords, there will be different filtrations depending on the construction of the chord chart, and not only two separated components that are gluing together at 100% of the filtration. Let us analyze each song.

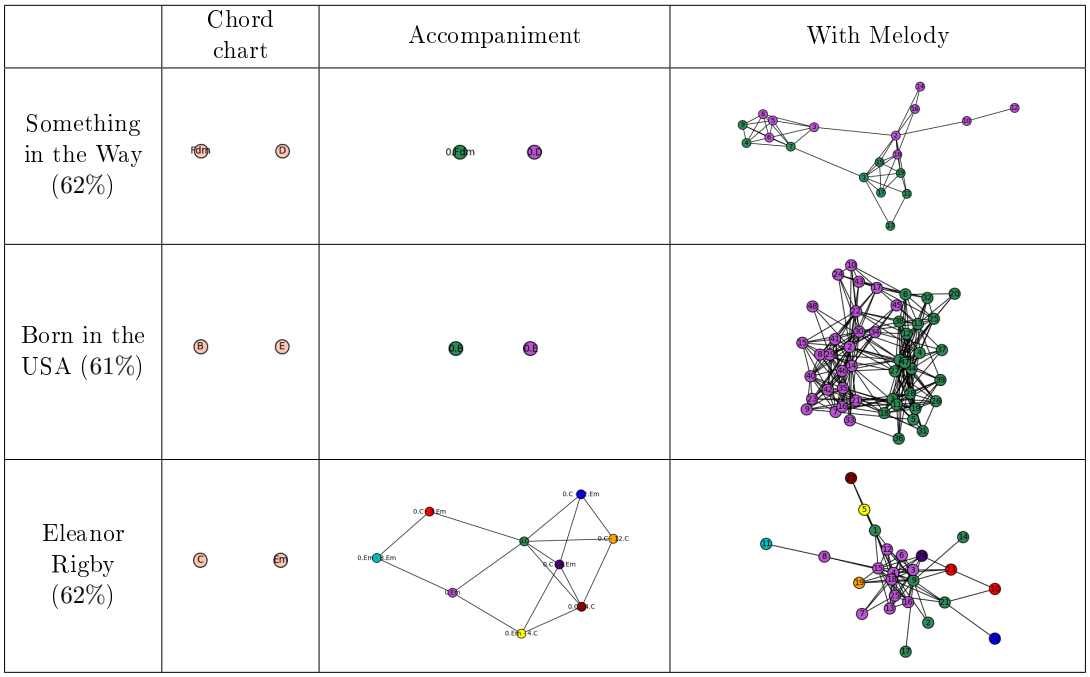


Figure 6.2.9: Comparison of the three songs *Something in the Way*, *Born in the USA* and *Eleanor Rigby* which are all built on two chords.

✻ **What's up.** This first song is based on one minor chord and two majors one. More precisely, D and Bm are linked by a relative transformation R, and we have:

$$R(D) = Bm \quad ; \quad LR(A) = D \quad ; \quad RLR(A) = R(D) = Bm$$

In consequence, the filtration at level 0 has three main scales of filtration, and barcodes is represented in Figure 6.2.10: at 69%, the edge $\{Bm, D\}$ appears, at 94% it is completed with the edge $\{A, D\}$, and finally the triangle is formed at 100% of the filtration. At level 1, we are in the basic case where there are only three vertices $\{0.A\}$, $\{0.D\}$ and $\{0.Bm\}$, so the family of barcodes and the filtration are the same.

On the opposite, we get fifty-nine distinct musical bars by adding the melody at level 2, so the melody seems to bring complexity to the song. The associated family of barcodes is given in Figure 6.2.11 and we can see that, if there are many bars in degree 0, there are not so many in degree 1. In addition they are quite short, especially compared to level 2 for *Born in the USA* in the previous analysis, where the melody also seems to add complexity to the song (the longest bar here lasts only 7% of the filtration, while the longest one for Bruce Springsteen's song lasts 20%). An illustration of the different levels of analysis for this filtration is given in Figure 6.2.12. At level 2, we still observe the property which says that the most colored vertex is the one associated with the tonality of the song: here it is the color $\{0.A\}$ that is the most represented, and A is the tonality of the song *What's up*.

✻ **I have a dream.** This second song is also based on three chords but, unlike the previous one, they are all major chords. In particular, that means that they are all connected by at least two successive PLR-transformations in $T[3, 4, 5]$: in fact, here we have $B\flat$, F and $E\flat$ and the following relations:

$$LR(E\flat) = B\flat = LR(F) \quad \text{and} \quad LRLR(E\flat) = F$$

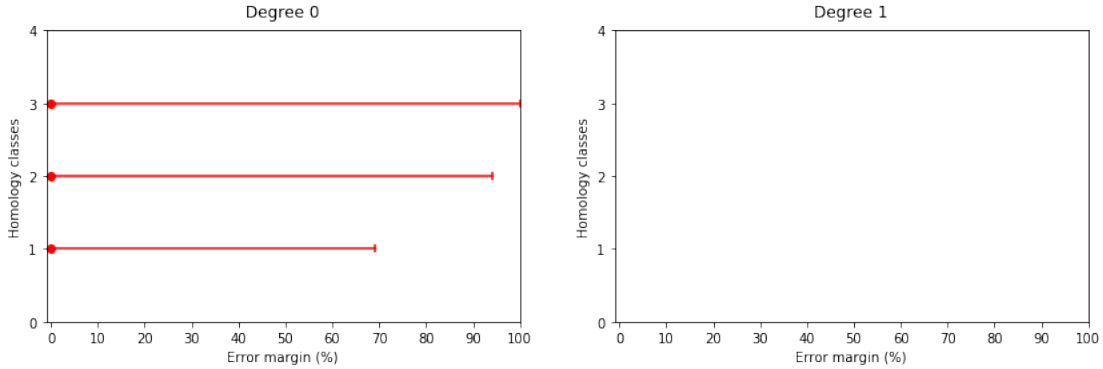


Figure 6.2.10: The associated family of barcodes for the harmonization of *What's up* at levels 0 and 1 of analysis.

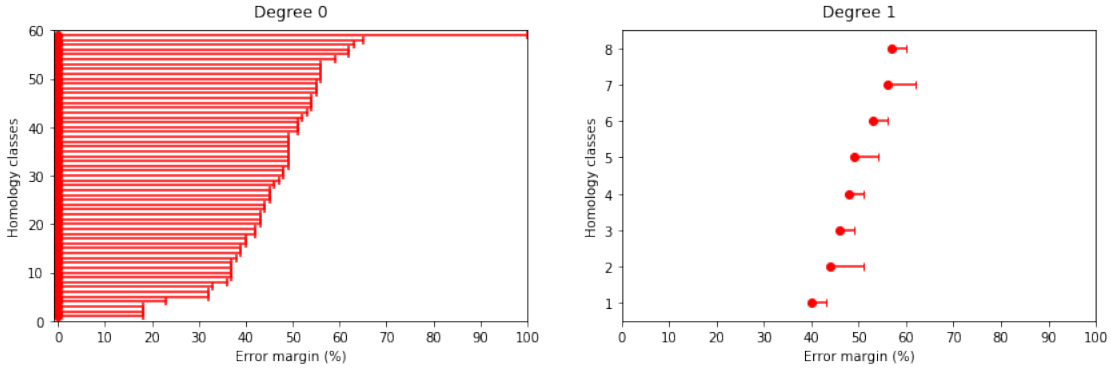


Figure 6.2.11: The associated family of barcodes for the harmonization of *What's up* at level 2 of analysis.

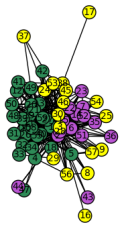
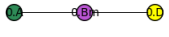
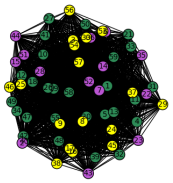
	Chord chart	Accompaniment	With Melody
65%			
94%			

Figure 6.2.12: Comparison of the three levels of analysis for *What's up* at a scale of 65% and 94%. We can observe at level 2 the clear separation between the three types of musical bars according to their accompaniment.

Notice that this progression is a classical one given by I-IV-V starting with $B\flat$. Thus, the complex at level 0 becomes connected at 87% where both edges $\{B\flat, F\}$ and $\{B\flat, E\flat\}$ appear together, and becomes a triangle at a scale of 100%. These moments of the filtration are visible on barcodes from Figure 6.2.13.

At level 1, we are in the same case than for *What's up*, which means that all the chords fall on the first position of the musical bar. In consequence, the filtrations are the same at level 0 and 1. The difference with the previous song comes with the melody: if both of them are built on three chords and level 1 does not allow to differentiate the songs, the addition of the melody (sixty-seven musical bars) at level 2 seems to reveal more complexity for *I have a dream*, which is quite visible with degree 1. In fact, the longest one-dimensional cycle lasts 19% of the filtration (see barcodes from Figure 6.2.14), while for *What's up* it was only 7%. For these two examples, there is no difference between the two first levels of analysis and the accompaniment, but the complexity of the melody is visible with level 2.

Finally, we give an illustration of the different filtrations for *What's up* respectively at a scale of 55% (the moment when level 2 becomes connected) and 87% (the same levels 0 and 1) on Figure 6.2.15: at level 2, we still have the same clustering between colored vertices according to their accompaniment.

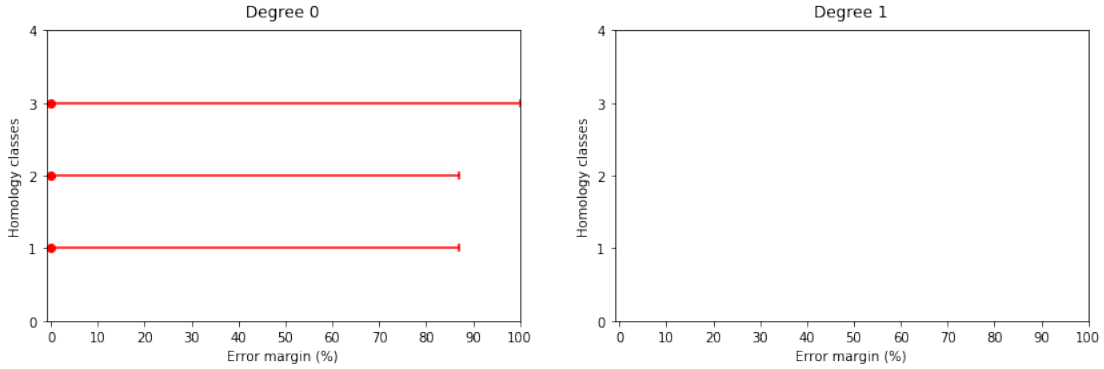


Figure 6.2.13: The associated family of barcodes for the harmonization of *I have a dream* at level 0 of analysis.

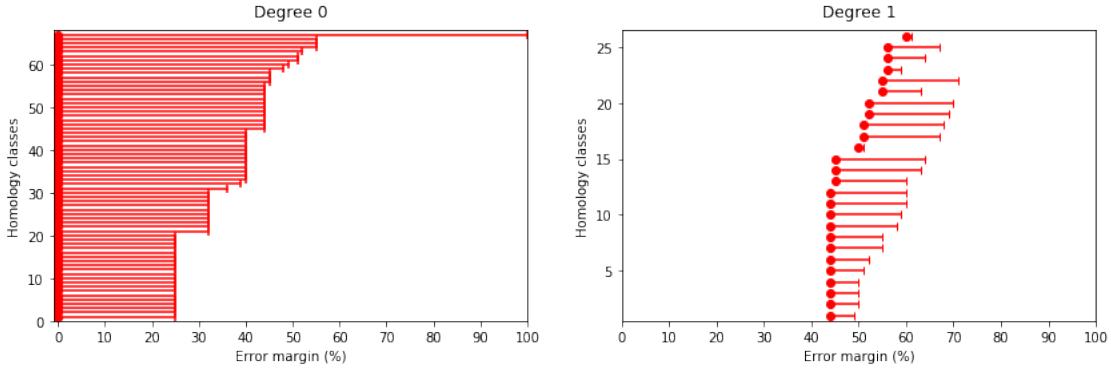


Figure 6.2.14: The associated family of barcodes for the harmonization of *I have a dream* at level 2 of analysis.

*** I love Rock N'Roll.** We finish this analysis with this three chords song that has an interesting level 1 of analysis. In fact, we have the same chord chart construction than for the previous one, that means construction I-IV-V starting with an E, so we have three major chords distant from transformations LR:

$$LR(B) = E = LR(A) \quad \text{and} \quad LRLR(B) = A$$

Thus, the associated family of barcodes at level 0 is the same than for *I have a dream* (Figure 6.2.13). For this example, the difference comes at level 1, where nine distinct musical bars

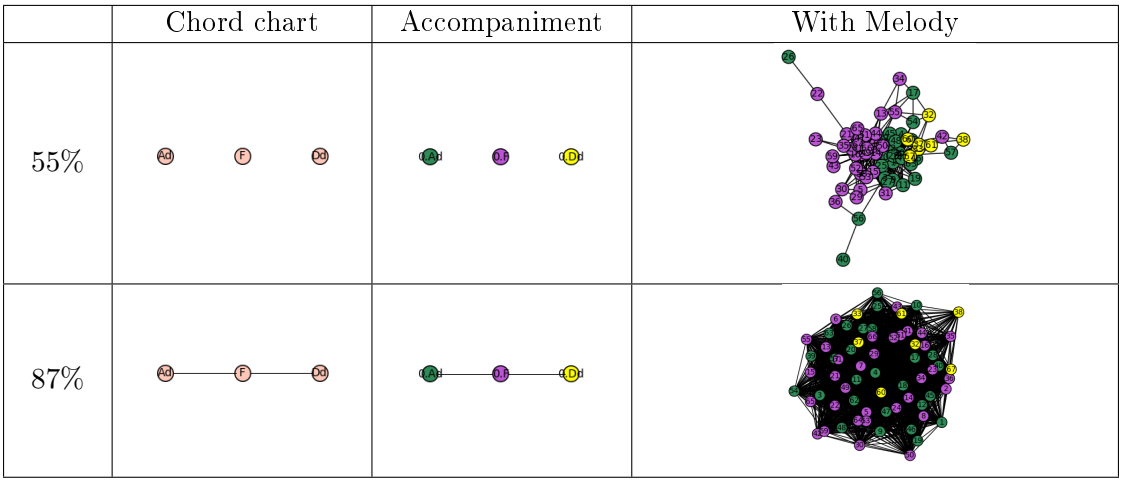


Figure 6.2.15: Comparison of the three levels of analysis for *I have a dream* at a scale of 55% and 87%. We can observe at level 2 the clear separation between the three types of musical bars according to their accompaniment.

appear. The family of barcodes is given in Figure 6.2.16: here the filtration becomes connected at a scale of 56% and it is also the moment when the longest one-dimensional cycle appears (it lasts precisely from 52% to 66%). This moment is represented in Figure 6.2.18: the musical bars are gathered by chords together with position and for instance, the generator in degree 1 is given by the five bars that start with $0.A$.

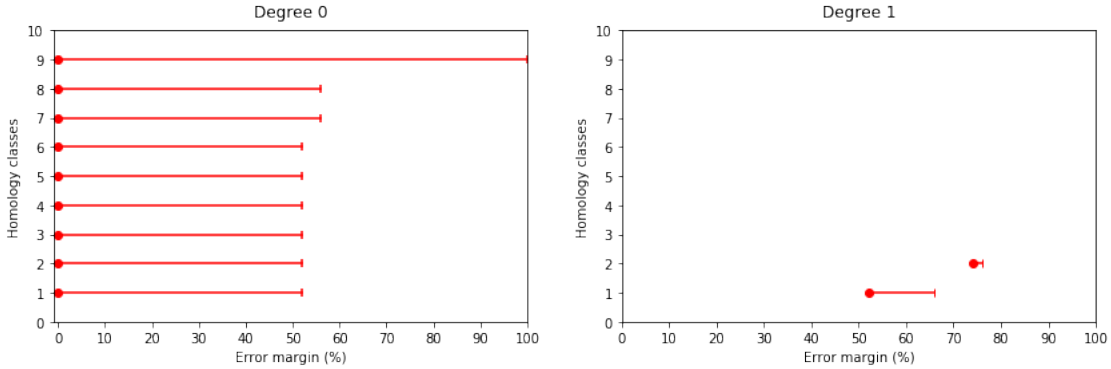


Figure 6.2.16: The associated family of barcodes for the harmonization of *I love Rock N'Roll* at level 1 of analysis.

At level 2, we have fifty-eight distinct musical bars and the barcodes are given in Figure 6.2.17. We still have a long generator in degree 1 that lasts 18% of the filtration (from 59% to 77%), and the complex at 70% seems to show this complexity. Moreover, the most represented color is the one that corresponds to the vertex $\{0.E\}$, and E is the tonality of this song.

✔ **Conclusion for 3-chords songs.** We saw here three different kinds of songs that are all built on only three chords and each level of analysis revels different musical features. Thus, we have a similar conclusion as for the 2-chords songs: for *What's up* and *I have a dream*, the two first levels of analysis are the same but the melody brings complexity for the second. In both cases, the musical bars are gathered by accompaniment. For *I love Rock N' Roll*, the filtration at level 1 reveals a much more complexity in the accompaniment than the two previous ones. This complexity is always visible by looking at barcodes in degree 1. The comparison of these three songs is illustrated in Figure 6.2.19.

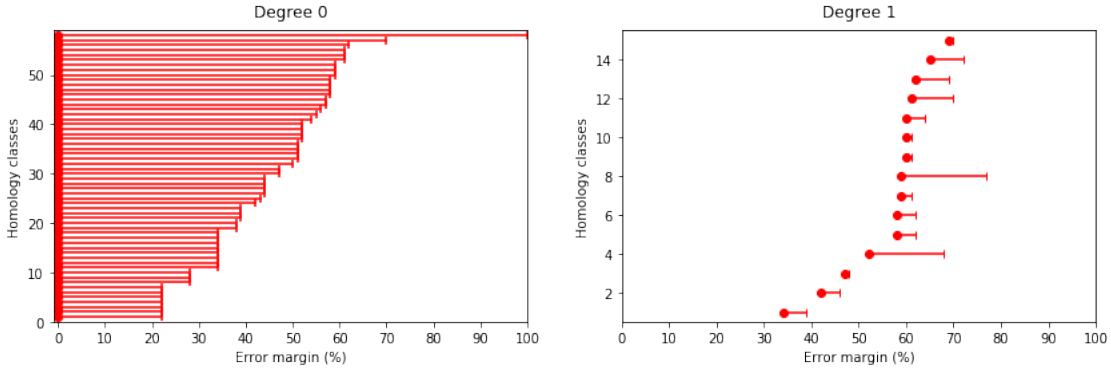


Figure 6.2.17: The associated family of barcodes for the harmonization of *I love Rock N'Roll* at level 2 of analysis.

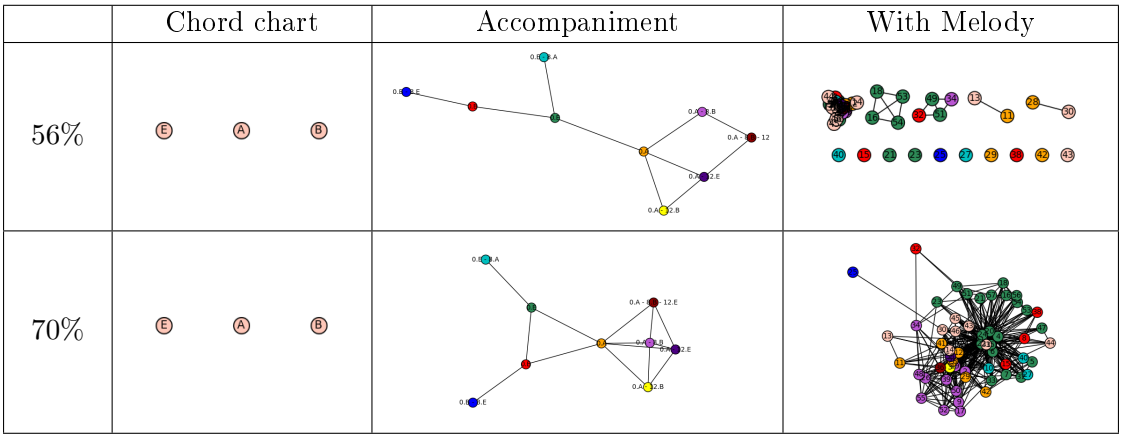


Figure 6.2.18: Comparison of the three levels of analysis for *I love Rock N'Roll* at a scale of 56% and 70%. We can observe at level 2 the clear separation between the three types of musical bars according to their accompaniment.

6.3. General case: songs with four chords

Now that we studied 2 and 3-chords songs, we can move to the most common case in Pop music, that means the one of songs that are built on four chords. More precisely, we will focus in a first place on the most famous harmonic progression of four chords, which is given by the following degrees progression:

I - V - vi - IV

where "I" is the tonic, "V" the dominant, "vi" the minor relative and "IV" the sub-dominant. For instance, starting with C-major chord as the tonic, we get the progression from Figure 6.3.1, and will call it the **Pop chord progression**. We will then study two other songs that have a different progression.

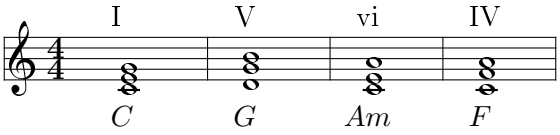


Figure 6.3.1: The classical Pop chord progression I-V-vi-IV starting with C major chord.

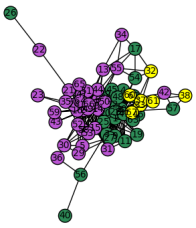
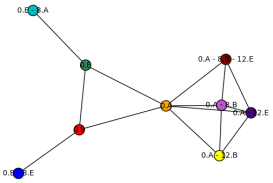
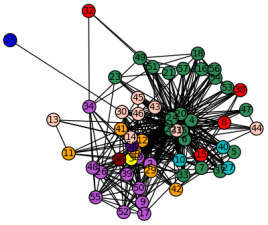
	Chord chart	Accompaniment	With Melody
What's up (65%)			
I have a dream (55%)			
I love Rock N' Roll (56%)			

Figure 6.2.19: Comparison of the three songs *What's up*, *I have a dream* and *I love Rock N' Roll* which are all built on three chords.

Notice that, since our DFT-distance is invariant by transposition and obviously does not care about the order of musical bars, all the songs that are built on the Pop chord progression will get the same filtration at level 0 (by relabelling the vertices). In particular, it will be interesting to see how level 1 allows us to differentiate songs with that same chords progression. More precisely, we will be able to find back the chord chart by looking at the filtration at level 0 in a first place and then, we will look at the accompaniment in itself by studying level 1 and compare different scores with the same harmonic progression. Finally, we will add the melody and see what it brings to the song. As for the 2 and 3-chords songs, we will see that, in despite of the same chord chart, 4-chords songs can also be differentiated and classified by **complexity**, using the comparison between these several levels of analysis.

For this study, we chose ten songs which are listed in Table 6.3.1. Notice that eight of them follow the I-V-vi-IV chord progression, while the last two have a different construction for comparison.

6.3.1. Pop chord progression I-V-vi-IV

Let us start with the classic harmonic progression I-V-vi-IV. We take as an example the song from U2, *With or Without You* (1987). This song has the Pop chord progression from Figure 6.3.1 but starting with a D:

$$D - A - Bm - G.$$

Notice that, in terms of PLR-transformation, we have the following relationship:

$$R(D) = Bm \quad ; \quad L(Bm) = G \quad ; \quad RL(D) = G \quad ; \quad RL(A) = D.$$

We computed the associated family of barcodes which is presented in Figure 6.3.2, and we also draw the whole associated filtration in Figure 6.3.3 (for level 0 of analysis). The barcodes and filtrations are the same for every 4-chords songs that are built on this progression (one just needs to rename the vertices).

Song	Artist	Year	Chord chart
Let it be	The Beatles	1970	$C - G - Am - F$
With or Without You	U2	1987	$A - D - Bm - G$
Zombie	The Cranberries	1994	$C - G - Em - D$
Happy Ending	Mika	2007	$C\sharp - G\sharp - A\sharp m - F\sharp$
Viva la Vida	Coldplay	2008	$G\sharp - D\sharp - Fm - C\sharp$
Danza Kuduro	Don Omar	2010	$C - G - Am - F$
Rude	Magic!	2014	$C\sharp - G\sharp - A\sharp m - F\sharp$
Despacito	Luis Fonsi	2016	$D - A - Bm - G$
We Found Love	Rihanna	2011	$F\sharp - D\sharp m - G\sharp m - B$
Get Lucky	Daft Punk	2013	$D - E - F\sharp m - Bm$

Table 6.3.1: The list of the ten studied 4-chords songs. The eight first follow the classical harmonic progression I-V-vi-IV, and the two others have a different construction for comparison.

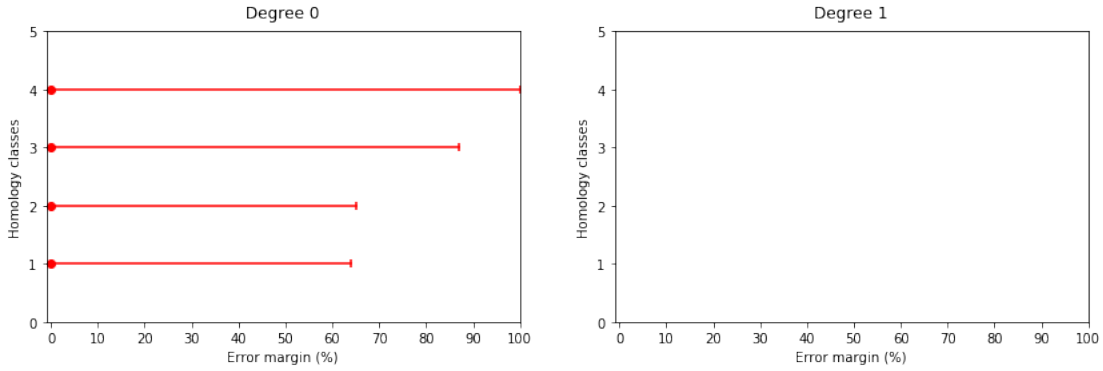


Figure 6.3.2: The associated family of barcodes with the Pop chord progression I-V-vi-IV.

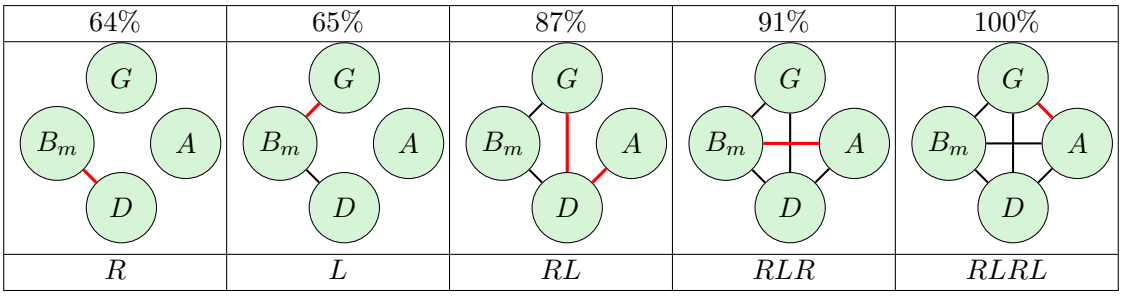
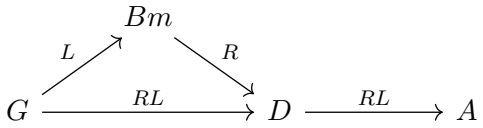


Figure 6.3.3: The filtration associated with the classical progression I-V-vi-IV starting with a D (for *With or Without You* in this example).

In degree 0, we have three main moments of filtration, which correspond respectively to transformations R (64%), L (65%) and composition LR (87%). In degree 1, there is no element of homology because of the triangle made of R, L and RL transformations that appears with a scaling parameter of 87%. Notice that all the complexes from Figure 6.3.3 respect the composition, which is consistent with the PLR-group and Cayley's graphs results from Theorem 5.2.6 (Section 5.2.2). For instance, at 87% of the filtration, we have the following diagram



We want now to look at the other levels of analysis. Here because we have eight songs to compare, we would not presented the associated family of barcodes for each songs as we did for 2 and 3-chords songs. We will simply illustrate the filtration of each song by displaying the first moment where it becomes connected, and thus for each moment of analysis. The table that summarized all these informations is presented in Figure 6.3.4. Level 0 is the same for each song because of the same chords progression, but we can easily observe the difference of complexity between levels 0 and 1: in fact, we have songs that are the same for both levels (*Viva la Vida*, *Zombie*) while some of them become much more complex (*Happy Ending*, *Rude*). In the same way, the complexity can be increased at level 2, like for *Viva la Vida* for example, while it seems more simple for other songs (*Rude*, *Zombie*). In Section 6.4, we will look for a way to quantify this notion of **complexity** and classify the different songs using this descriptor. We will also try to refine the graph we choose as a representative of the song.

6.3.2. Other 4-chords songs progressions

In this paragraph, we focus on the two 4-chords songs from Table 6.3.1 that do not follow the classical Pop chord progression I-V-vi-IV. These songs are *Get Lucky* from the French band Daft Punk (2013) and *We Found Love* from the American singer Rihanna (2011). We will present filtrations and barcodes in both cases, and compare them with the previous results.

❋ **Get Lucky.** This song is based on the 4-chords $\{D, E, F^\sharp m, Bm\}$, which is not a progression I-V-vi-IV. Actually, we have the following relationships between chords:

$$R(D) = Bm, \quad L(D) = F^\sharp m, \quad RL(F^\sharp) = Bm, \quad PRL(Bm) = E$$

$$LRPRL(E) = F^\sharp m \quad \text{and} \quad RPRL(D) = E.$$

Notice that using the transformation P provides more complicated relationships than the simple Pop progression. Using this, we build the filtration from Figure 6.3.5 at level 0 of analysis. In this illustration, we see that the filtration is closed from the one associated with the progression I-V-vi-IV, mainly because we have exactly the same shapes from 87% to the end. The most important change is at 82%, the first moment when the filtration is connected: in fact, here the graph is a tree that looks like a simple line

$$\{F^\sharp m, D\} + \{D, Bm\} + \{Bm, E\}$$

which is something we never get with the previous progression.

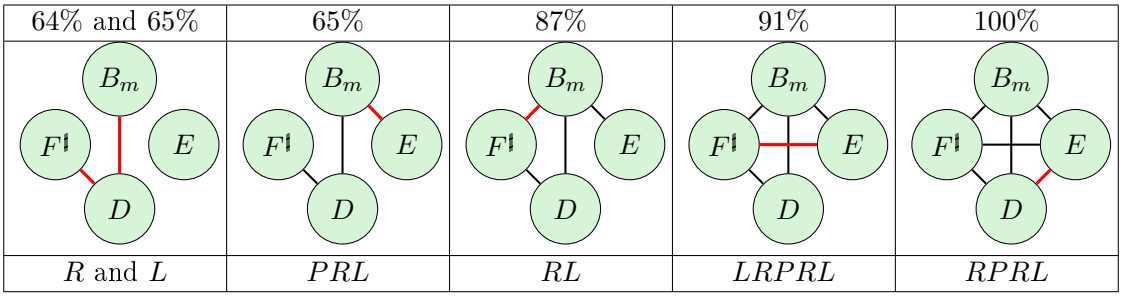


Figure 6.3.5: The filtration associated with *Get Lucky* at level 0 of analysis.

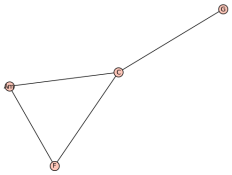
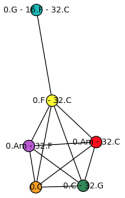
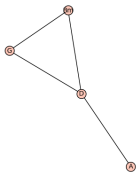
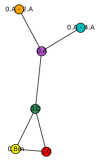
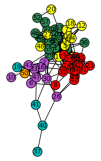
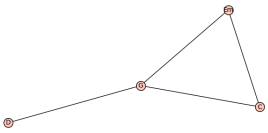
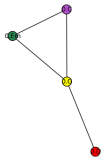
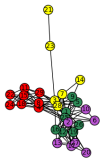
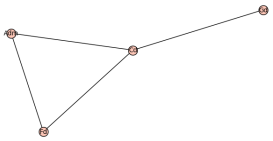
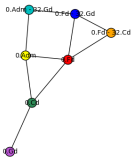
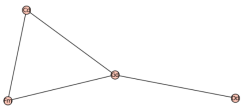
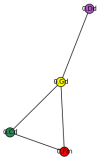
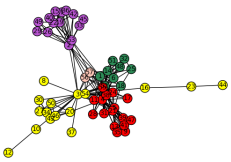
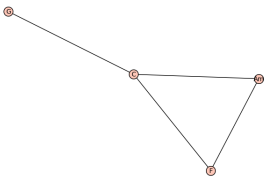
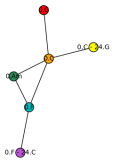
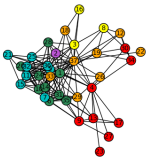
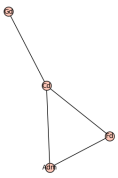
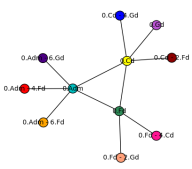
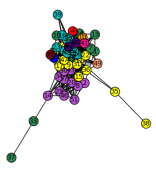
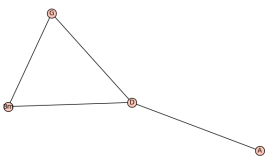
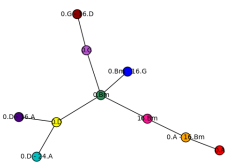
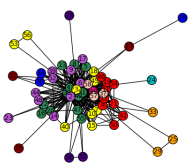
	Chord chart	Accompaniment	With Melody
Let it be			
	87%	94%	65%
With or Without You			
	87%	66%	61%
Zombie			
	87%	87%	65%
Happy Ending			
	87%	63%	69%
Viva la Vida			
	87%	87%	58%
Danza Kuduro			
	87%	68%	64%
Rude			
	87%	63%	63%
Despacito			
	87%	59%	64%

Figure 6.3.4: Comparison of the 4-chords songs that are built on the Pop chord progression I-V-vi-IV. For each level, we have chosen the first moment when the filtration becomes connected.

At level 1, we have the same filtration, since each chord is played at the first position of the musical bar. At level 2, musical bars are also gathered by accompaniment. We represented the three levels of analysis in Figure 6.3.6 (according to the first moment where it is connected). Notice that, for this last level, the complex has only twenty-six vertices so it looks quite light.

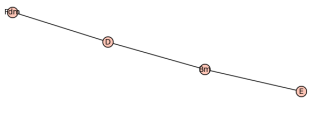
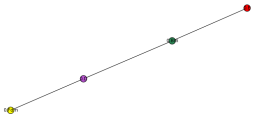
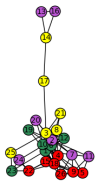
	Chord chart	Accompaniment	With Melody
Get Lucky			
	82%	82%	61%

Figure 6.3.6: The associated graphs for *Get Lucky* for each level of analysis. The chosen scale for each graph is the one that corresponds to the first moment when the filtration becomes connected.

❋ **We Found Love.** This second song is based on four chords $\{F^\sharp, D^\sharp m, B, G^\sharp m\}$, which is again not a progression I-V-vi-IV. Unlike *Get Lucky*, the progression here is more simple in terms of PLR-transformations:

$$R(F^\sharp) = D^\sharp m, \quad R(B) = G^\sharp m, \quad L(B) = D^\sharp m$$

$$RL(B) = F^\sharp, \quad RL(G^\sharp m) = D^\sharp m, \quad \text{and} \quad RLR(G^\sharp m) = F^\sharp.$$

In particular, the filtration becomes connected at 71%, with the basic line given by successive transformations R, L and R:

$$\{F^\sharp, D^\sharp m\} + \{D^\sharp m, B\} + \{B, G^\sharp m\}.$$

As for *Get Lucky*, we recover a simple tree, but this time not the same complexes as for the classical Pop progression. We display the general filtration at this level of analysis in Figure 6.3.7. Also notice that, unlike all the previous filtrations, we only have four main times of filtrations before it is fully connected: 69%, 70% and 94%.

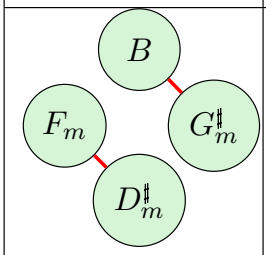
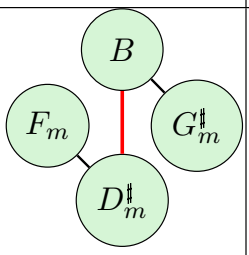
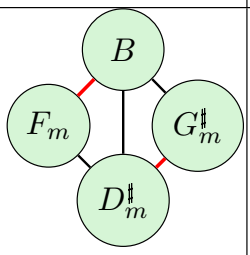
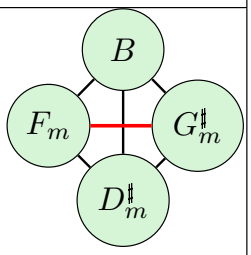
69%	70%	94%	100%
			
R	L	RL	RLR

Figure 6.3.7: The filtration associated with *We Found Love* at level 0 of analysis.

If the chord chart looks more simple with this progression, we will get with that song a general accompaniment that looks much more complicated. This is visible on Figure 6.3.8, where we compare the different level of analysis according to the first moment where the filtration becomes connected. With only four chords at the basis, we get eight distinct musical bars for

the accompaniment. At level 2, we have forty-two vertices, and notice that the musical bars are gathered by accompaniment. Moreover, the complex we represented for level 1 shows vertices that are grouped around the vertex $\{0.F\sharp\}$, with a direct relative edge $\{0.F\sharp, 0.D\sharp m\}$, which provides the tonality: actually, $R(F\sharp) = D\sharp m$, and the most colored vertex at level 2 is the one associated with musical bar $\{0.D\sharp m\}$, which is the tonality of *We Found Love*.

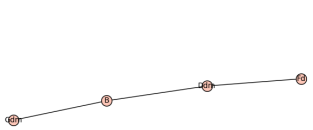
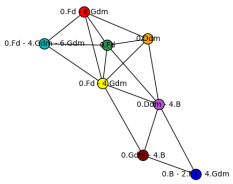
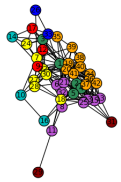
	Chord chart	Accompaniment	With Melody
We Found Love			
	70%	79%	56%

Figure 6.3.8: The associated graphs for *We Found Love* for each level of analysis. The chosen scale is the one that corresponds to the first moment when the filtration becomes connected.

We see with this example that, if this song and *Get Lucky* have the same shape at level 0 as a first connected moment (a tree in a direct line), the complexity of the general accompaniment differs and level 1 allows us to see this difference. The same remark goes for the melody, and this difference of complexity is also visible on barcodes at degree 1, which we represented in Figure 6.3.9 for both songs: in fact, for *We Found Love*, we have much more bars which are also quite long compare to *Get Lucky*.

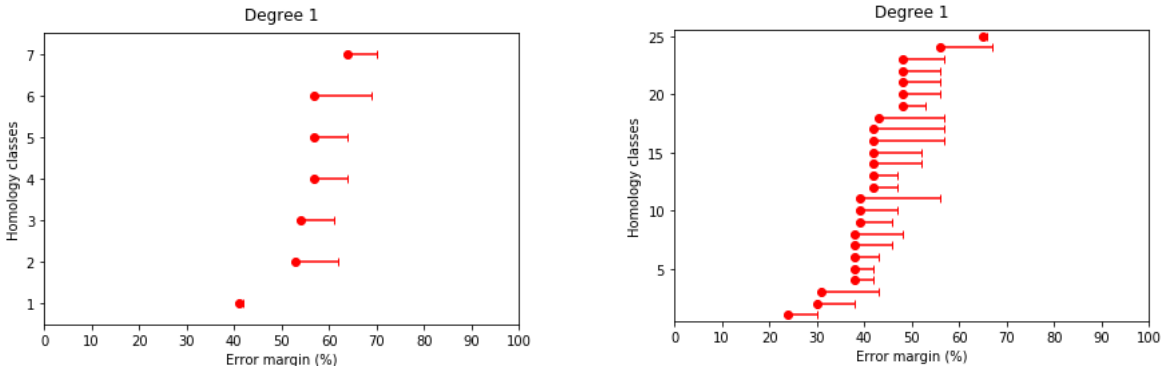


Figure 6.3.9: Comparison of barcodes in degree 1 for *Get Lucky* (left) and *We Found Love* (right) at level 2 of analysis.

✔ **Conclusion for 4-chords songs.** The 4-chords analysis confirms the borderline cases in the sense that several songs with exactly the same chord chart (up to transpositions) can have less or more complicated accompaniment and melody. The comparison between level 0 and 1 underlines these differences. Moreover, level 1 shows that we can have a complicated accompaniment, and the addition of the melody can reduce this complexity or, on the contrary, increase this complexity. This is also true for songs that do not follow the same 4-chords progression. Our purpose now is to to quantify this complexity and, in this perspective, we need to decide precisely in a canonical way which representative complex (or graph) we choose to represent a given song.

6.4. A graph-type for any musical score

6.4.1. The general construction

In the previous sections, we studied different Pop songs based on two, three and four chords. For each song and each level of analysis, we arbitrarily choose to display the first moment when the filtration becomes connected and simply displayed this graph as a kind of representative of the song. The idea now is to refine this moment and create an algorithm that provides a representative **graph-type** for any of these studied Pop songs, and which can next be extended to any music piece.

First notice that we commonly use either the word "graph" or "complex" to designate the illustration of a song at a given scale. In fact, the representations we show in this paper are always given by graphs, for the simple reason that filled triangles and higher dimensional simplices would overload the drawing, and thus prevent the hand study. Also recall that we build our filtrations with the Vietoris-Rips method, which is the most common practice in Topological Data Analysis. From Remark 4.1.2, we know that a complex obtained in this way (at a given scale) is fully determined by its 1-skeleton, since a collection of vertices is a simplex if and only if they form a complete subgraph, so there is no loss of information in focusing on the 1-skeleton instead of the whole complex. For example, the Vietoris-Rips complex in two dimensions is given by the graph with filled triangles.

Of course, the homology of the complex depends on the higher simplices, and one must decide whether this is the most relevant calculation to perform. In many applications of TDA, the point cloud is made of points sampled from a larger space, very close together, and one does not expect significant cycles of short length, but rather large, "macroscopic" ones. In particular, filling every triangle seems natural, when computing the homology.

As we have realized when studying 2, 3 and 4-chords songs, the 1-skeleton is often an interesting object in and of itself, and we have started to discuss the graphs and not the complexes. Even the 3-cycles seem to convey information that should not be discarded: for instance, if we just focus on the classic Pop progression of four chords I-V-vi-IV we previously studied, there is no homology in degree 1 for the first level of analysis. Indeed, there are four vertices and no time to have the outline of the square before the diagonals appear. The result is that we can not quantify the one-dimensional homology in this case other than by zero, which is not satisfying because of the presence of loops. Moreover, for *Get Lucky* or *We Found Love*, which do not follow this same progression of chords, we would also have to quantify the one-dimensional homology by zero while the filtrations are definitely not the same.

Thus, we have elected to study more closely the 1-skeleton, to be called simply the **graph** of the point cloud, and to compute its homology (still in degrees 0 and 1), rather than that of the full complex. As an example, we computed the associated family of barcodes associated with graphs for filtration I-V-vi-IV, and the result is given in Figure 6.4.1.

In this illustration, degree 0 is the same than the original construction, but the barcode in degree 1 has changed: in fact, at the first moment when the filtration is connected (87%), instead of considering that the homology is zero, the triangle

$$\{D, G\} + \{G, Bm\} + \{Bm, D\}$$

provides a one-dimensional cycle (illustrated in this example with the progression $D-A-Bm-G$ from *With or Without You*, see Figure 6.3.3). Morever, this cycle represents a bar that lasts until the end of the filtration. In the same way, at a scale of 91%, there is a second triangle

$$\{D, A\} + \{A, Bm\} + \{Bm, D\}$$

that appears, and together with the first, they give the outline of the square. The last edge $\{D, E\}$ complete the graph and provides the last element of homology in degree 1, which is here simply a point (a bar that starts at 100% of the filtration).

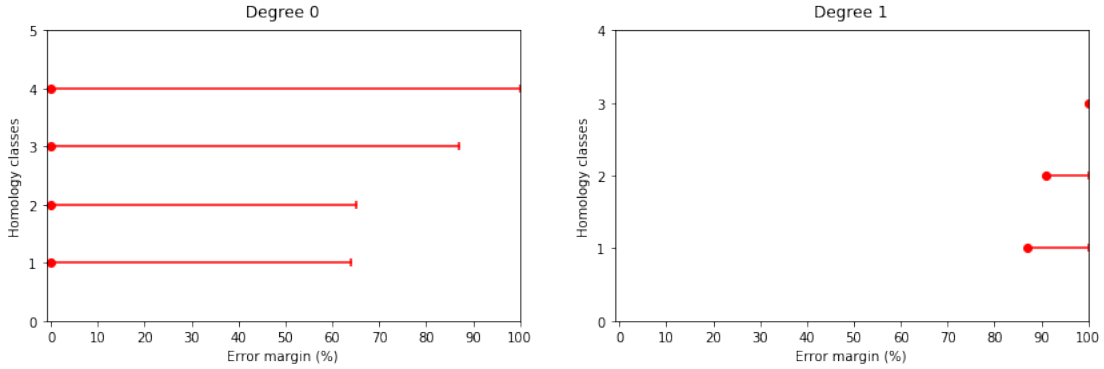


Figure 6.4.1: The associated family of barcodes with the classical progression I-V-vi-IV by looking at the filtration made of graphs (level 0 of analysis).

For comparison, we chose to display the associated family of barcodes with filtration for *We Found Love* at level 0, i.e. for the chord chart $F\sharp -D\sharp m, B, G\sharp m$. The results are presented in Figure 6.4.2: we can see that now the filtration are well differentiated with these new barcodes. In fact, homology in degree 1 are not the same any more, and we see for instance the apparition of two triangles at 94% that are not present for I-V-vi-IV progression. This new analysis will thus allows us to better compare filtrations and also complexity for several songs.

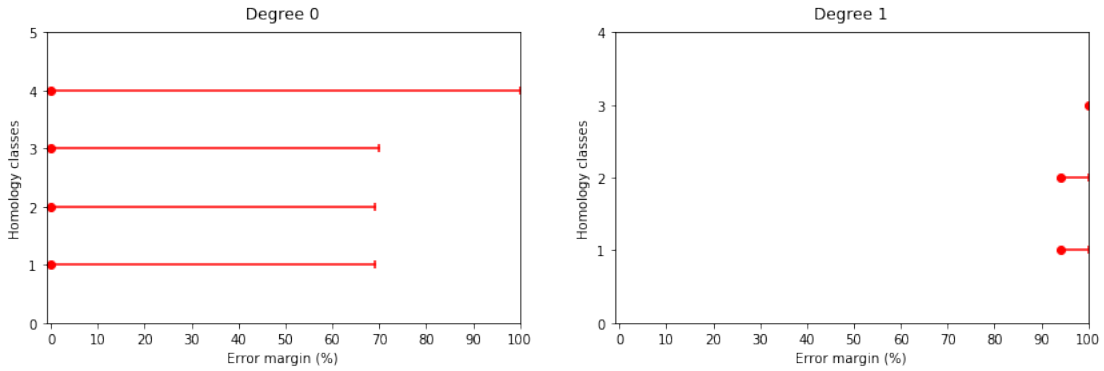


Figure 6.4.2: The associated family of barcodes with the song *We Found Love* by looking at the filtration made of graphs (level 0 of analysis).

We now want to extract just one graph from this construction, called the **graph-type**. First, we can notice the particular shape of the barcodes and particularly the complementarity in the progression of the bars between the two degrees: in degree 0, each bar starts at 0% and successively stops before 100%, whereas in degree 1 they all end at 100% and start after 0%. More precisely, all the bars in degree 1 start after the first moment when a first connection appears in degree 0, and we will use this property to refine the scale we choose to study for our filtrations. In fact, the graph-type should be defined using the main ideas of persistent homology, that means looking at the length of the bars and more precisely the ones that persist. The algorithm we propose to compute the graph-type is the following one:

✻ **Graph-type Protocol.** Let $\{\mathrm{BC}_0(\mathcal{S}), \mathrm{BC}_1(\mathcal{S})\}$ be the family of barcodes (in degrees 0 and 1) associated with a score $\mathcal{S}$. Each one of the two barcodes defines a natural subdvision $s_0 = 0, s_1, s_2, \ldots, s_k = 100$ of the interval $[0, 100]$, where the s_i's are the integers where some bar either starts or ends (or put differently, the complex stays the same when the scaling parameter goes through $]s_i, s_{i+1}[$).

i) In degree 0, define $t_0 > 0$ to be the integer s_i such that $|s_{i+1} - s_i|$ is maximal (the longest interval on which the graph remains the same, as far as homology in degree 0 can see).

ii) In degree 1, take $t_1 > t_0$ to be the integer s_i such that $|s_{i+1} - s_i|$ is maximal (the longest interval on which the graph remains the same, as far as homology in degree 1 can see). If there is only one bar that starts at 100%, take $t_1 = t_0$.

iii) Take the complex associated with filtration at a scale of t_1%. The associated 1-skeleton is called the **graph-type** of the filtration.

Remark 6.4.1. For the scale t_0, we voluntarily take $t_0 > 0$ since the very first interval $[s_0, s_1]$ is usually the longest, while the graph is obviously not interesting at these scales yet. We could also choose to define t_1 in the same way as we defined t_0, but there would be no guarantee that $t_1 > t_0$ (think of a complex which starts with a large number of disjoint triangles, which are then connected by segments not creating new cycles, so that H_0 decreases as H_1 stays constant). However, the rationale is that small values of the scaling parameter should be ignored until the homology in degree 0 "stabilises", and then homology in degree 1 should be used to select the most significant interval, presumably discarding large values of the scaling parameter. Thus, we compute t_1 as t_0 but with the constraint that $t_1 > t_0$, which is why we truncate $\mathrm{BC}_1(\mathcal{S})$.

Remark 6.4.2. Note that with this construction it is always possible to return to the initial filtration of complexes instead of the graphs by simply computing the associated clique-complex for each graph. Note also that the graph we get from this protocol can be non connected.

We will now end this section by computing the associated graph-type with the previous protocol for the classical Pop chord progression I-V-vi-IV.

Example 6.4.3. Recall the classical progression I-V-vi-IV, from which for instance the chord chart of *With or Without You* is built on. We are going to construct the associated graph-type for level 0. The filtration was given in Figure 6.3.3 and associated family of barcodes in 6.4.1.

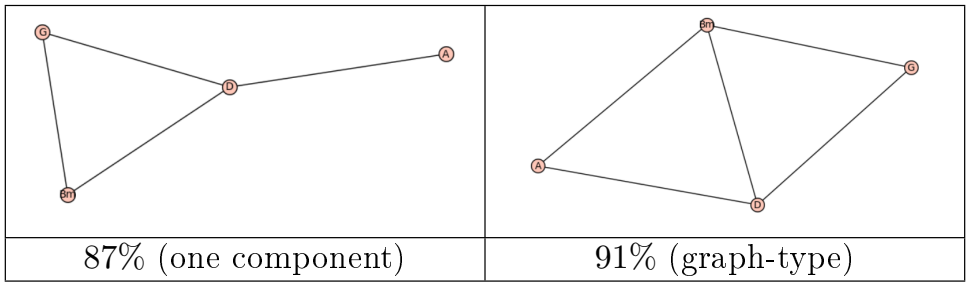
Let us apply our algorithm. In degree 0, the first bar stops at 64% of the filtration, and then the three left bars stop respectively at 65%, 87% and 100%. Thus, by computing the successive differences, we have

$$65\% - 64\% = 1\%, \quad 87\% - 65\% = 22\% \quad \text{and} \quad 100\% - 87\% = 13\%$$

so the longest bar lasts 22% and starts at 65%. Therefore, we have $t_0 = 65\%$. Then, we truncate the barcode in degree 1 at 65%, which does not change anything in that case because the first bar starts at 87%, and we have three different bars that start respectively at 87%, 91% and 100%. By computing again successive differences, we get that

$$91\% - 87\% = 4\% \quad \text{and} \quad 100\% - 91\% = 9\%$$

so the longest bar lasts 9% and starts at 91%, and thus $t_1 = 91\%$. Finally, the graph-type is given by the filtration at a scale of 91%. We display the result graph-type for this example in Figure 6.4.3, and compare it with the moment we choose in a first place (the first moment where the filtration becomes connected). Here we refined the chosen scale by taking into account both degrees 0 and 1, and the persistent interpretation of the the barcodes in terms of length of the bars.



87% (one component)	91% (graph-type)

Figure 6.4.3: The graph-type associated with the classical progression I-V-vi-IV (here precisely for *With or Without You*) which is given at a scale of 91% (right) compare with the first moment when the filtration is connected (at 87%, left).

As a musical interpretation of this first example, we can observe that the graph with a scaling parameter of 91% is given by a square that is cut by an edge in the middle: this edge gives the tonality of the song, to the relative minor. In above Example 6.4.3, this edge is given by $\{D, Bm\}$ so we know that the song is either in D major or in Bm minor key, and finally the mode can be find with the color graphs by comparing level 1 with level 2, as we already saw. The graph-type for this progression seems therefore to be a good representative of the score.

We thus created an algorithm that provides an associated graph-type for any musical piece, and we saw with the example of progression I-V-vi-IV that it fixes the scaling parameter we chose in a first place. Moreover, it takes into account the degree 1 that was not present because of the filled triangles in the Vietoris-Rips filtration, so it seems to be a more precise approach. We already used the filtrations to compare songs in terms of complexity, now we are going to use these particular graphs to quantify this notion. We will apply this analysis in the case of 5 and 6-chords songs (see Section 6.5), and we will also propose a direct application of graph-types to scales in Tonnetze in the next paragraph.

The construction of an algorithm that takes a MIDI file and gives a corresponding graph by analyzing persistent homology (in degree 0 and 1) is a satisfying result in itself.

6.4.2. Digression: graph-types for scales and Tonnetze

To support our definition of a **graph-type** associated with a musical score, we applied this process to musical objects from Chapter 5, that is, scales and Tonnetze. The results are presented in Figure 6.4.4. We recover the different shapes we studied for the different musical objects: for example, we have the circle of thirds for the major and minor scales, and the torus for different two-dimensional Tonnetze $T[a, b, c]$. This seems to confirm that we are indeed associating a musically consistent graph (or complex, by taking the corresponding clique-complex) with our musical objects, and this supports our algorithm.

6.4.3. How to measure the complexity of a song

In the previous section, we give an algorithm that takes a musical piece and returns the associated graph-type. We already compared graphs together in the study of 2, 3 and 4-chords and tried to measure the complexity of the different songs. We will now quantify this notion using this particular graph-type and its homology.

The process is quite simple: in order to continue to base our work on homology, we will just measure the complexity of a graph by looking at the number of its one-dimensional loops, that means taking a basis for H_1. Thus, by computing the Betti numbers $\beta_1 = \dim H_1$, we obtain a way to compare several graphs together.

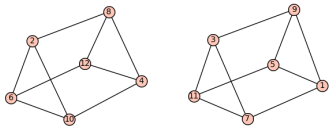
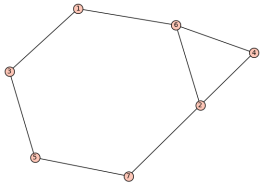
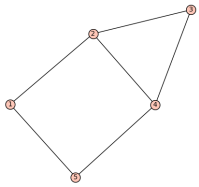
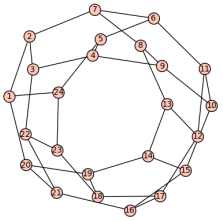
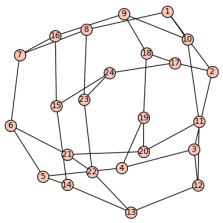
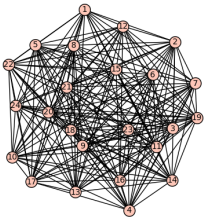
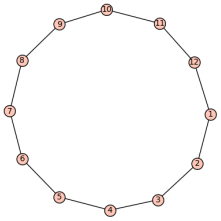
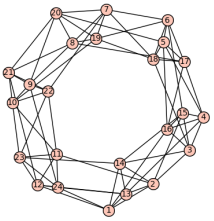
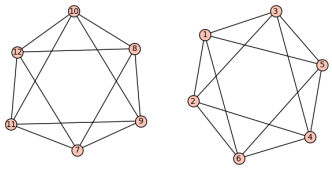
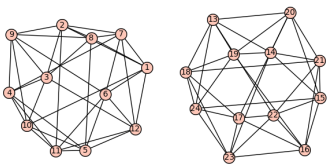
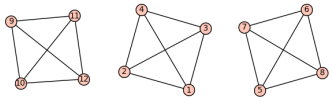
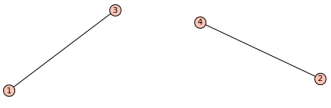
	
Chromatic scale (91%)	Diatonic Major/minor scale (95%)
	
Pentatonic Major/minor scale (98%)	$T[1,2,9]$ and $T[2,3,7]$ (56%)
	
$T[1,3,8]$ and $T[3,4,5]$ (56%)	$T[1,4,7]$ (89%)
	
$T[1,1,10]$ and $T[2,5,5]$ (64%)	$T[1,5,6]$ (58%)
	
$T[2,2,8]$ (68%)	$T[2,4,6]$ (57%)
	
$T[3,3,6]$ (65%)	$T[4,4,4]$ (82%)

Figure 6.4.4: The associated graph-types for some musical objects (scales and Tonnetze).

Definition 6.4.4. Let $\mathcal{S}_{\mathfrak{P}}$ be a score that corresponds to a given musical piece $\mathfrak{P}$. By applying the graph-type protocol, we can associate a graph $\mathcal{G}_{\mathcal{S}_{\mathfrak{P}}}$ to $\mathcal{S}_{\mathfrak{P}}$. The **complexity $\mathcal{C}_{\mathcal{S}_{\mathfrak{P}}}$ of the score** $\mathcal{S}_{\mathfrak{P}}$ is then defined by

$$\mathcal{C}_{\mathcal{S}_{\mathfrak{P}}} = \beta_1\big(\mathcal{G}_{\mathcal{S}_{\mathfrak{P}}}\big) = \dim H_1\big(\mathcal{G}_{\mathcal{S}_{\mathfrak{P}}}\big).$$

As an example, we produced the associated graph-type for some 4-chords songs of Table 6.3.1 at level 1 (accompaniment) and computed the associated Betti numbers β_1. The results are presented in Figure 6.4.5.

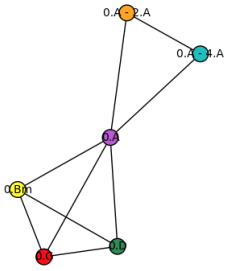
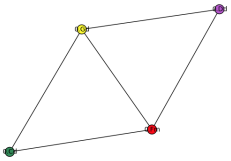
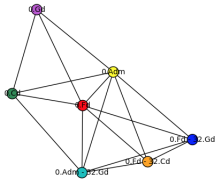
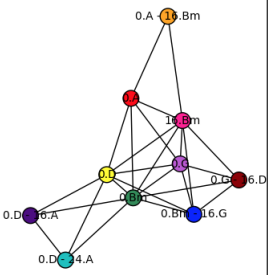
With or Without you (76%)	Viva La Vida (91%)	Happy Ending (76%)	Despacito (76%)
			
$\beta_1 = 4$	$\beta_1 = 2$	$\beta_1 = 10$	$\beta_1 = 16$

Figure 6.4.5: The associated graph-types and Betti numbers β_1 for different 4-chords songs at level 1 of analysis (accompaniment).

We can now compare these different graphs by their complexity: we confirm that *Viva la Vida* has the simplest accompaniment (same as at level 0), while in this case *Despacito* has the most complex one. For these examples, we do not have a lot of vertices so the complexity simply confirms what we can observe on the graphs, but it allows us to have a better visualization for more complicated cases. For instance, we compare the complexity of the graph-types for levels 0, 1 and 2 in the case of 4-chords songs, and we display these informations in an histogram given in Figure 6.4.6. Each graph of this figure represents one level of analysis, and we simply normalized the Betti numbers to quantify the complexity as a percentage.

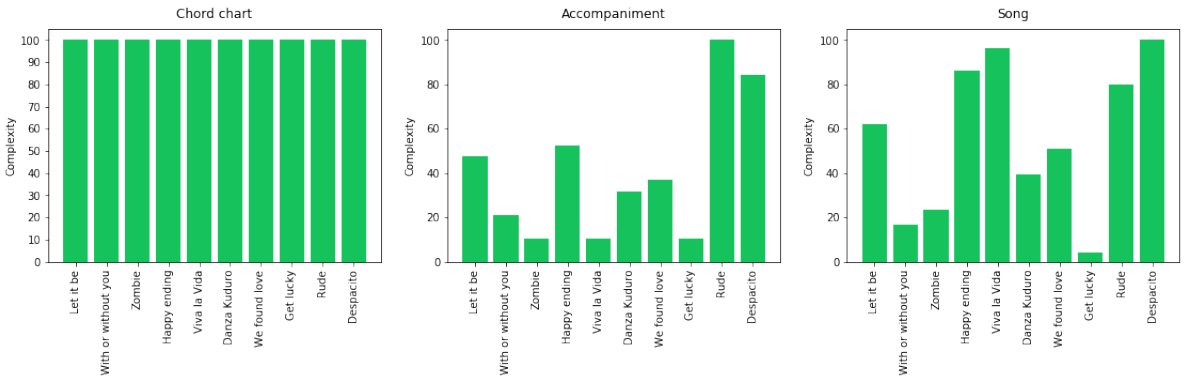


Figure 6.4.6: The complexity of the ten studied 4-chords songs for each level of analysis.

The complexity at level 0 is the same for each piece because the associated graph-type is always the same (the square with the diagonal). At level 1, we have differences that appear, and the most *complex* song is *Rude* ($\beta_1 = 19\%$), while the less are *Zombie*, *Viva la Vida* and *Get Lucky*, which are the songs that have the same filtration for chord chart and accompaniment. By

adding the melody at level 2, we have songs that earn complexity as *Viva la Vida* for instance, which was already what we observed by looking at graphs with the previous hand analysis.

Remark 6.4.5. Notice that our construction did not allow us to differentiate the chords chart, especially between *We Found Love*, *Get Lucky* and the progression I-V-vi-IV. This is due to the fact that we decided to look at the graph-type using both 0 and 1 degrees, and here the filtrations are different before the bars appear in degree 1. One idea might be to adjust the graph-type even more precisely by looking at the time that the filtration spends not having one-dimensional cycle. However, the graph-types we propose are a compromise that seems to have a consistent musical interpretation, considering the previous results. In any case, it is always possible to go back to the initial filtration to analyze the score in depth and compare other features.

6.5. To go further: songs with five and six chords

To conclude with this section devoted to the problem of harmonization songs, we simply give an application of what we have done for songs with five and six chords. In these cases, it becomes much more complicated to analyze the results by hand, firstly because we have less canonical choice for the chord chart (most of the time in the studied songs, the base will be the famous progression I-V-vi-IV, but with one or two additional chords that can be really different). Secondly, the number of vertices can increase quickly at level 1 (in some cases, we have about thirty distinct musical bars), so the comparison can be difficult. We will thus use the graph-types as a representation of the songs for each level, and then the complexity to compare them together.

For doing the analysis, we took ten songs built on five chords and ten others based on six, from different artists or groups and different generations of the Pop category (from the '80s to nowadays). The lists of the pieces are presented in Table 6.5.1. For each song, we represented the complexities for all songs in Figures 6.5.1 and 6.5.1, and we also computed the associated family of graph-types, shown in Figure 6.5.3 and 6.5.4.

	Song	Artist	Year	Chord chart
5-chords	Eye of the Tiger	Survivor	1982	$B\flat - Fm - A\flat - Cm - E\flat$
	Sweet Dreams	Eurythmic	1983	$Cm - G\sharp - G - F - Fm$
	Forever Young	Alphaville	1984	$C - G - Am - F - Dm$
	Hallelujah	Rufus Wainwright	2001	$C - Am - F - G - Em$
	The Scientist	Coldplay	2002	$D - E\flat - F - Dm - C$
	Bad Romance	Lady Gaga	2009	$F - G - Am - C - E$
	Skyscraper	Demi Lovato	2011	$G - C - Em - D - Am$
	Let Her Go	Passenger	2012	$C - D - Em - G - B$
	Wake Me Up	Avicci	2014	$Bm - G - D - E\flat$
	Hello	Adele	2014	$Fm - A\flat - E\flat - D\flat - Cm$
6-chords	Imagine	John Lennon	1971	$C - F - Am - Dm - G - E$
	You're the one that I want	J. Travolta, O. Newton-John	1978	$Am - G - F - C - Em - Em$
	Hold the Line	Toto	1978	$Bm - E - F\sharp m - D - C\sharp - Dm$
	Wind of Change	Scorpions	1990	$F - Dm - Am - G - C - E$
	Wonderwall	Oasis	1995	$F\sharp m - A - E - B - D - G$
	It's My Life	Bon Jovi	2000	$Cm - A\flat - E\flat - B\flat - C - F$
	Complicated	Avril Lavigne	2002	$Dm - B\flat - F - C - Dm - Gm$
	Party in the USA	Miley Cyrus	2009	$F\sharp - D\sharp m - C\sharp - A\sharp m - G\sharp m - B$
	Someone Like You	Adele	2001	$A - C\sharp m - F\sharp m - D - E - Bm$
	The Sound of Silence	Disturbed	2015	$F\sharp m - E - D - A - Bm - F\sharp$

Table 6.5.1: The list of the twenty studied 5-chords and 6-chords songs.

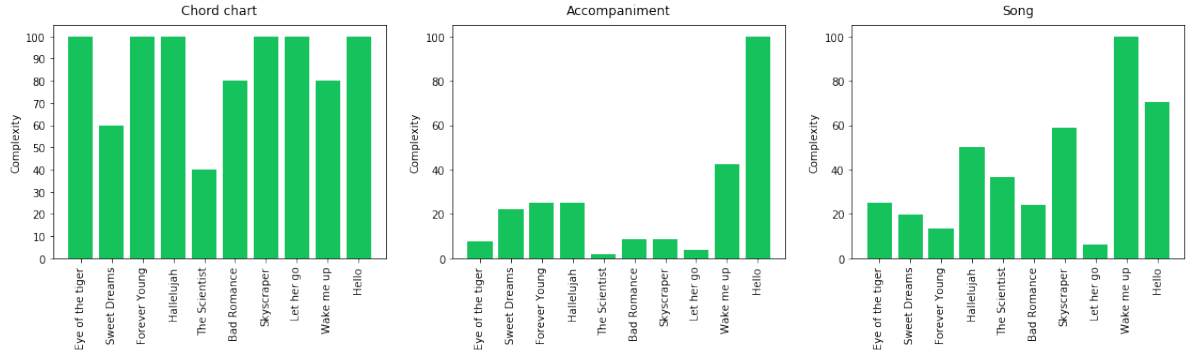


Figure 6.5.1: The complexity of the ten studied 5-chords songs for each level of analysis.

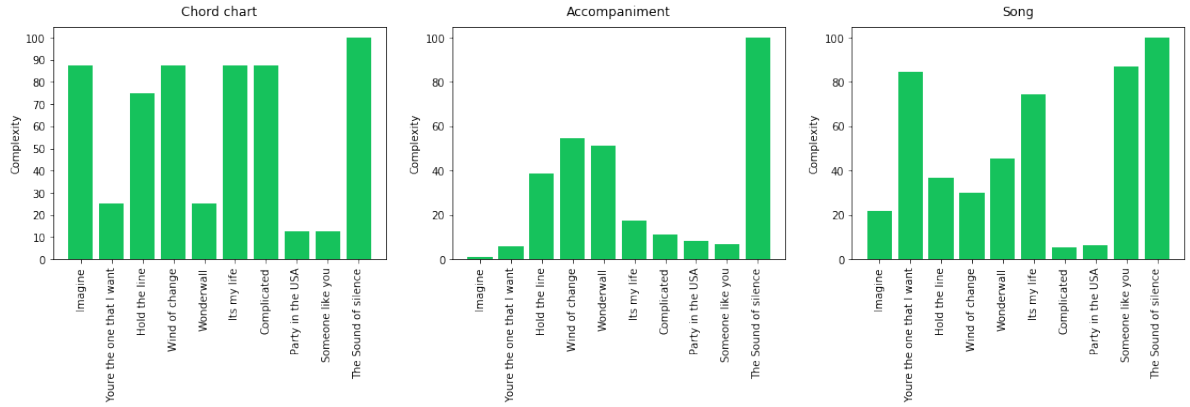


Figure 6.5.2: The complexity of the ten studied 6-chords songs for each level of analysis.

Let us briefly analyze the results for the songs with five chords: the first histogram shows that more than half of the songs have a complexity of 100%, and these are the songs whose graph type looks like a diamond. At level 1, only one song keeps this maximum value (*Hello*), and we see that its graph-type seems to be the most complicated one, with a large number of vertices and also a scale close to 100%, so the graph is almost complete. On the contrary, some songs like *Let Her Go* get a very simple accompaniment, which can be seen in the graph-type. Moreover, the song with the lowest values for its chord chart and accompaniment (*The Scientist*) is the only one with an non-connected graph. Note also that its graph-type is almost complete at level 2, and its complexity is a bit higher (which means that the melody adds complexity to the song). On the contrary, the song *Wake me up* has a quite high complexity at level 0 (80%), the second highest at level 1 (43%) and the higher for the whole song. Therefore, we have songs with constant complexity and others that gain or lose complexity with the level of analysis.

For songs with six chords, we have almost the same results: for example, the song with the highest value at each level is *The Sound of Silence*, and it is quite clear considering the successive graph-types. On the other hand, the song *Party in the USA* has almost the lowest value for each level, and the graphs are really light. Note that the graph-type at level 0 for this song and for *Someone like you* looks the same (same chord chart up to a transposition), and the graph represents a loop of 6 chords, which shows that there is a particular choice of chords in the Euler's Tonnetz (simply a direct line that loops after a while). There are also some songs that can have a low complexity for the chord chart and accompaniment, but a high value for the song itself, as is the case for *You're the One that I want* or *Someone Like You*. Again, some songs have varying complexity depending on the level of analysis, while others are fairly constant, and our approach provides a way to compare and analyze this.

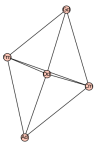
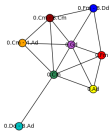
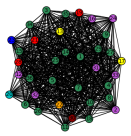
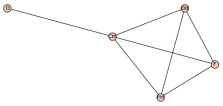
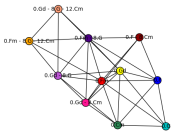
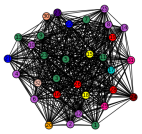
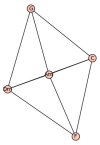
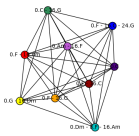
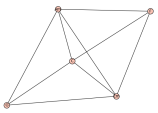
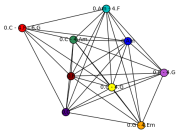
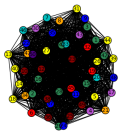
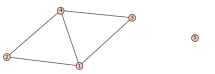
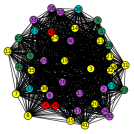
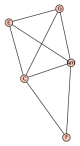
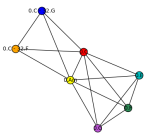
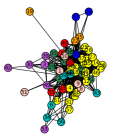
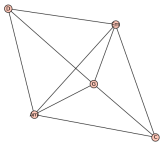
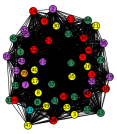
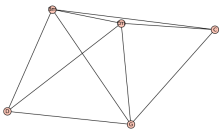
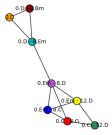
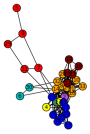
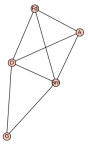
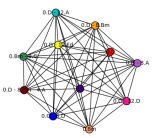
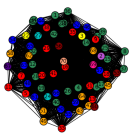
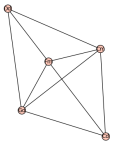
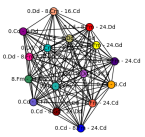
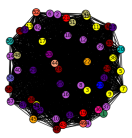
	Chord chart	Accompaniment	With Melody
Eye of the Tiger			
	91%	77%	95%
Sweet Dreams			
	83%	69%	97%
Forever Young			
	91%	88%	53%
Hallelujah			
	93%	91%	94%
The Scientist			
	64%	80%	95%
Bad Romance			
	88%	86%	69%
Skyscraper			
	91%	88%	97%
Let Her Go			
	93%	69%	67%
Wake Me Up			
	88%	88%	97%
Hello			
	93%	96%	97%

Figure 6.5.3: Comparison of the ten studied 5-chords songs from Table 6.5.1 by their graph-type.

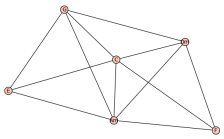
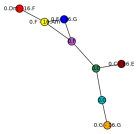
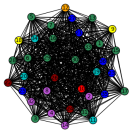
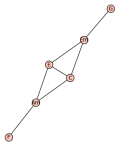
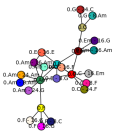
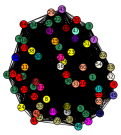
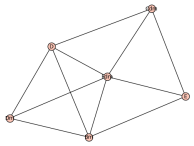
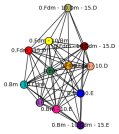
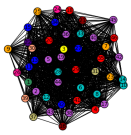
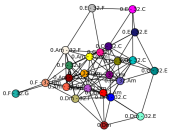
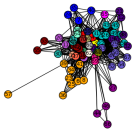
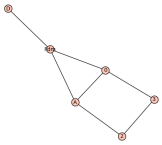
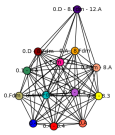
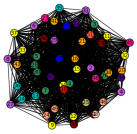
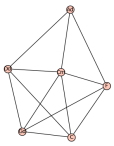
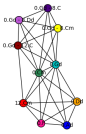
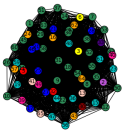
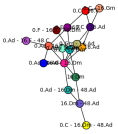
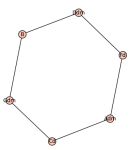
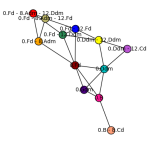
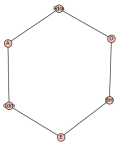
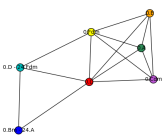
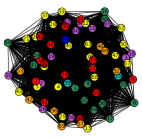
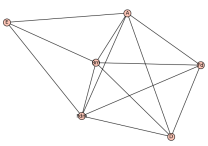
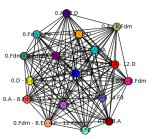
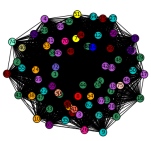
	Chord chart	Accompaniment	With Melody
Imagine			
	87%	66%	97%
You're the one that I want			
	73%	53%	97%
Hold the line			
	89%	87%	97%
Wind of Change			
	87%	63%	66%
Wonderwall			
	75%	86%	95%
It's My Life			
	91%	76%	96%
Complicated			
	91%	58%	56%
Party in the USA			
	81%	60%	64%
Someone like you			
	93%	78%	96%
The Sound of Silence			
	91%	90%	98%

Figure 6.5.4: Comparison of the ten studied 6-chords songs from Table 6.5.1 by their graph-type.

Chapter 7.

Classification of musical style

In this section, we aim to apply our DFT-approach to the well-known problem of automatic classification of musical style. In fact, one of the main goals of applying persistent homology and Topological Data Analysis to music is to provide some algorithms that would be able to "recognize" the style of a given music piece by analyzing the associated family of barcodes. There is already some work on this topic, and we can cite the famous article by Bergomi [10] which is a precursor of the subject. In this chapter, we propose a new way of approaching automatic style analysis by combining the DFT together with persistent homology.

We will start by detailing the strategy, which consists of transforming a barcode into a family of points in $\mathbb{N}^2$, and computing statistical features on the length of the bars (mean, standard deviation and entropy), as suggested in [39]. Therefore, we will select several MIDI files from of different musical styles, starting from Heavy Metal to Baroque, and compare them by clustering in $\mathbb{R}^3$. All the musical data we are going to use are listed in a database which is available on the following dedicated web page: `https://math-musique.pages.math.unistra.fr/midi.html`.

7.1. Statistics on barcodes

First of all, we need to clarify what we mean by comparing barcodes. In fact, it seems that there are several ways to analyze a barcode, and we propose here a statistical approach. We assume that any musical piece $\mathfrak{P}$ can be associated with a score $\mathcal{S}_{\mathfrak{P}}$ and thus a family of barcodes $\big(\mathrm{BC}_d(\mathcal{S}_{\mathfrak{P}})\big)_{d\geq 0}$ by applying the DFT-method, as described in Part II of this manuscript. As with the previous musical applications, we will essentially focus on $d \in \{0,1\}$ as a starting point, leaving higher dimensions for future work. We start by turning any barcode $\mathrm{BC}_d(\mathcal{S}_{\mathfrak{P}})$ into a family of points in $\mathbb{N}^2$, following the definition below.

Definition 7.1.1. Let $\mathcal{S}_{\mathfrak{P}} = \{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$ be a score associated with a music piece $\mathfrak{P}$ containing N distinct musical bars. Let $\big(\mathrm{BC}_d(\mathcal{S}_{\mathfrak{P}})\big)_{d\geq 0}$ be the corresponding family of barcodes. We denote by r_d the number of bars for the barcode $\mathrm{BC}_d(\mathcal{S}_{\mathfrak{P}})$. For each degree d, an element of the $\mathrm{BC}_d(\mathcal{S}_{\mathfrak{P}})$ is a generator of the associated homology group H_d, so it can be seen as a pair of element

$$(b_i, d_i) \in \mathbb{N}^2$$

where b_i and d_i are respectively referring to the birth and the death time of the corresponding *ith* homology generator. Note that in our case d_i is always finite since we are only dealing with finite filtration. Therefore, each $\mathrm{BC}_d(\mathcal{S}_{\mathfrak{P}})$ is as a subset of $\mathbb{N}^2$:

$$\mathrm{BC}_d(\mathcal{S}_{\mathfrak{P}}) = \{(b_i, d_i) \mid 1 \leq i \leq r_d, \text{ with } (b_i, d_i) \in \mathbb{N}^2\}$$

Now each musical score $\mathcal{S}_{\mathfrak{P}}$ has a corresponding family of barcodes $\big(\mathrm{BC}_d(\mathcal{S}_{\mathfrak{P}})\big)_{d\geq 0}$ and, for each $d \geq 0$, $\mathrm{BC}_d(\mathcal{S}_{\mathfrak{P}}) \subset \mathbb{N}^2$. We can thus compute some basic statistics on each barcode, by considering the length of the bars.

Definitions 7.1.2. Let $\mathcal{S}_{\mathfrak{P}} = \{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$ be a score associated with a music piece $\mathfrak{P}$ containing N distinct musical bars, and let $\big(\mathrm{BC}_d(\mathcal{S}_{\mathfrak{P}})\big)_{d\geq 0}$ be the corresponding family of barcodes. We denote by r_d the number of bars in degree d. We define three statistical values on the length of the bars for $\mathrm{BC}_d(\mathcal{S}_{\mathfrak{P}})$:

1. The d-**persistent mean value** associated with $\mathcal{S}_{\mathfrak{P}}$ is given by

$$\mu_d(\mathcal{S}_{\mathfrak{P}}) = \frac{1}{r_d}\sum_{i=1}^{r_d}(d_i - b_i)$$

2. The d-**persistent standard deviation** associated with $\mathcal{S}_{\mathfrak{P}}$ is given by

$$\sigma_d(\mathcal{S}_{\mathfrak{P}}) = \sqrt{\frac{1}{r_d}\sum_{i=1}^{r_d}\big((d_i - b_i) - \mu_d(\mathcal{S}_{\mathfrak{P}})\big)^2}$$

3. The d-**persistent entropy** associated with $\mathcal{S}_{\mathfrak{P}}$ is given by

$$\epsilon_d(\mathcal{S}_{\mathfrak{P}}) = -\sum_{i=1}^{r_d}\rho_i \log \rho_i$$

where the coefficients ρ_i are defined by $\rho_i = \frac{d_i - b_i}{\sum_{i=1}^{r_d} d_i - b_i}$.

At this point, we want to compare several musical pieces together, and we are now able to transform each associated musical score into a family of points in $\mathbb{R}^3$. In fact, through our DFT-approach, a musical piece corresponds to a family of barcodes which is itself a subset in $\mathbb{N}^2$, that can be summarized by three coordinates: its mean, its standard deviation and its entropy. Thus, for a fixed degree d, we can represent a given family of musical pieces in $\mathbb{R}^3$ by displaying the corresponding barcodes using these statistics. A natural analysis would then be to observe the distribution of the points, potential clusters, and then conclude whether or not our DFT-approach is able to classify different musical styles. The general process we will follow throughout this chapter is summarized in Figure 7.1.1. For computational constraints and as previously announced, we will only work with the barcodes in degrees 0 and 1, and especially in degree 0, which seems to provide the most consistent results so far.

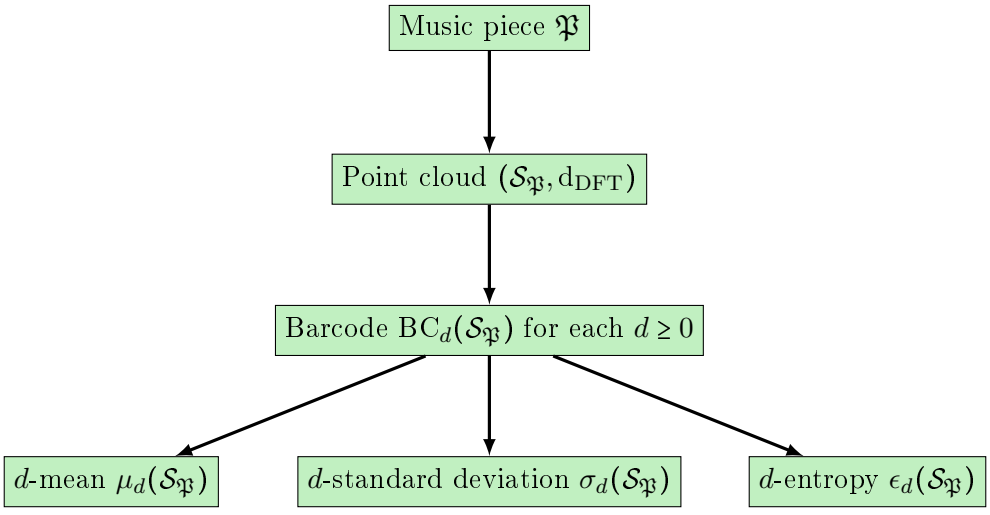


Figure 7.1.1: The transition from a music piece $\mathfrak{P}$ into a point $\big(\mu_d(\mathcal{S}_{\mathfrak{P}}), \sigma_d(\mathcal{S}_{\mathfrak{P}}), \epsilon_d(\mathcal{S}_{\mathfrak{P}})\big) \in \mathbb{R}^3$.

7.2. Comparison between Heavy Metal and Classical music

For this first analysis, we deliberately chose the most obvious possible comparison: Heavy Metal compare to Classical music. Before doing some sharp analysis, our first goal is to provide a way to confirm our approach: indeed, if it turns out that the results do not separate these two styles that seem different in many ways, then we would have to make some modifications. Otherwise it would be a first validation of the process.

In all this section, we are going to oppose Heavy Metal style against different types of Classical music: Baroque, Classical and Romantic. The idea is to take different bands from Heavy Metal against one representative of these different styles of Classical musical, which will be musical pieces by J.S. Bach, W.A. Mozart and F. Chopin, respectively. In that purpose, we have chosen five bands that are representative of the Metal style: Metallica, Iron Maiden, Judas Priest, Scorpions and Nightwish. Each group represents a different way of approaching the Heavy Metal style, and here is a short presentation of the bands:

- **Metallica** is an American band formed in 1981 and specialized in **Thrash Metal**.
- **Iron Maiden** is an English band formed in 1975 which is representative of the **new wave of British Heavy Metal**.
- **Judas Priest** is an English band formed in 1969 and specialized in **traditional Heavy Metal**.
- **Scorpions** is a German band formed in 1965 and specialized in **Hard Rock**.
- **Nightwish** is a Finnish band formed in 1996 which is representative of the **Symphonic Metal**.

For each band, we took fifteen songs from their discography, making sure that the data covered as much of the band's run as possible. The used data are presented in Table 7.2.1.

Heavy Metal				
Metallica	**Iron Maiden**	**Judas Priest**	**Scorpions**	**Nigthwish**
Bleeding Me (1996)	Aces High (1984)	Beyond the Realms of Death (1978)	Always somewhere (1979)	Amaranth (2007)
Creeping death (1984)	Blood Brothers (2000)	Breaking the Law (1980))	Bad boys running wild (1984)	Bless the Child (2002)
Disposable Heros (1986)	Dance of Death (2003)	Diamonds and Rust (1977)	Believe in love (1988)	Bye Bye Beautiful (2007)
Enter Sandman (1991)	Die with your boots on (2015)	Electric Eye (1982)	Big city nights (1984)	Elan (2015)
Fade to Black (1984)	Fear of the Dark (1992)	Halls of Valhalla (1982)	Blackout (1982)	End of All Hope (2002)
Fuel (1997)	Hallowed be thy Name (1982)	Hell Bent for Leather (1978)	Catch your train (1976)	I Want My Tears Back (2011)
Hero of the day (1996)	Phantom of the Opera (1980)	Hell Patrol (1990)	No one like you (1982)	Last Ride of the Day (2011)
Master of Puppets (1986)	Run to the Hills (1982)	Living after Midnight (1980)	Passion rules the game (1988)	Nemo (2004)
Nothing Else Matters (1991)	The Book of Souls (2015)	Metal Gods (1980)	Rock you like a Hurricane (1984)	Over the Hills and Far Away (2001)
Orion (1986)	The Clansman (1998)	Nightcrawler (1990)	Rhythm of love (1988)	She is My Sin (2000)
Sanitarium (1986)	The Evil that Men Do (1988)	Painkiller (1990)	Send me an Angel (1990)	Sleeping Sun (1998)
Seek and destroy (1982)	The Number of the Beats (1982)	The Sentinel (1983)	Still Loving You (1984)	Storytime (2011)
St Anger (2003)	The Trooper (1983)	Touch of Evil (1986)	Tease me please me (1990)	The Islander (2007)
The call of Ktulu (1984)	Total Eclipse (1982)	Turbo Lover (1976)	Wind of Change (1990)	The Phantom of the Opera (2002)
The memory remains (1997)	Wasted Years (1986)	Victim of Changes (1976)	When the smoke is going down (1982)	Wish I Had an Angel (2004)

Table 7.2.1: The five representative Heavy Metal bands and the fifteen corresponding songs used in the analysis.

7.2.1. Baroque: Johann Sebastian Bach

Let us begin with the first comparison: Baroque and Heavy Metal. The Baroque music was composed around 1600 to 1750 and represents a transition between the Renaissance and the Classical periods. One of the main characteristics of the Baroque music is the *basso continuo*, which is a bass line played throughout a piece by one or more low instruments such as a cello,

a viola or a double bass, and is built around one or two soloists. The main instruments of the Baroque music are the recorder, the viola, the harpsichord, the lute and the organ.

As a representative of the Baroque style, we choose to focus on Johann Sebastian Bach (1685-1750), and more specifically on three different styles of composition from his repertoire: the Chorals, the Preludes and the Fugues. The data used for the analysis are presented in Table 7.2.2 (fifteen pieces for each style).

Johann Sebastian Bach								
Chorals			Preludes			Fugues		
No. 1 BWV 269	No 6. BWV 281	No 11. BWV 41	No. 1 BWV 870	No. 7 BWV 876	No. 14 BWV 883	No. 1 BWV 870	No. 7 BWV 876	No. 14 BWV 883
No. 2 BWV 347	No 7. BWV 389	No 12. BWV 65	No. 3 BWV 872	No. 8 BWV 877	No. 16 BWV 885	No. 3 BWV 872	No. 8 BWV 877	No. 16 BWV 885
No. 3 BWV 2	No 8. BWV 40	No. 13 BWV 261	No. 4 BWV 873	No. 9 BWV 878	No. 17 BWV 886	No. 4 BWV 873	No. 9 BWV 878	No. 17 BWV 886
No 4. BWV 9	No 9. BWV 248	No. 14 BWV 184	No. 5 BWV 874	No. 10 BWV 879	No. 21 BWV 890	No. 5 BWV 874	No. 10 BWV 879	No. 21 BWV 890
No 5. BWV 9	No 10. BWV 687	No. 15 BWV 277	No. 6 BWV 875	No. 11 BWV 880	No. 23 BWV 892	No. 6 BWV 875	No. 11 BWV 880	No. 23 BWV 892

Table 7.2.2: The forty-five pieces from Chorals, Preludes and Fugues from Johann Sebastian Bach repertory that are used for analyzing the Baroque style.

We have the working data: each music piece has an associated MIDI file and thus a score. Then, following the process that is presented in Figure 7.1.1 and Definition 7.1.1, for any degree $d \in \{0, 1\}$, each musical piece is transformed into a barcode in $\mathbb{N}^2$, and then into a point in $\mathbb{R}^3$, characterized by three coordinates: its mean value, its standard deviation and its entropy. We then display all the points and compare the distribution and possible clusters. The first results of the comparison between the Heavy Metal and the Baroque style are shown in Figure 7.2.1, where each line corresponds to a Heavy Metal band and each column represents a degree (0 or 1).

This figure shows immediate clusters at degree 0: indeed, there is a clear separation between each one of the representative bands of Heavy Metal and all the pieces of Bach, which are on the contrary clustered together. On the other hand, degree 1 does not seem to reveal such information: in fact, we have much fewer points, especially for the Chorals, which means that there are not always generators of homology in degree 1 for these pieces. One explication could be that the pieces we took from Bach, and especially the Chorals, have on average less distinct musical bars than the Heavy Metal ones (about twenty for the Chorals compared to more than a hundred for the pieces from Metallica).

We therefore focus more specifically on degree 0, which we detailed in Figure 7.2.2: in particular, we highlighted the clusters by showing the projection on each axis and adding the convex hull for each piece of Bach and each band of Heavy Metal. We thus see an interesting phenomenon appearing: there is indeed a clear separation between each band of Heavy Metal and Bach, but this separation is the same for each group considered. For instance, let us look at the first projection on the (mean, standard deviation)-axis: for Metallica, it is clear that each piece has a mean of less than 50, while Chorals, Fugues and Preludes from Bach have a mean greater than this value. This is a first result, that needs to be understood in itself. It is interesting to note that for Iron Maiden, Judas Priest, Scorpions and Nightwish, we have the same separation on the first projection, with a persistent mean around 50. In the same way, the two left projections show a clear separation when the entropy is around 5 or 6, and this is the same separation for all the bands analyzed. More precisely, each Heavy Metal song from each band seems to be in the same window, delimited by a mean value less than 50, a standard deviation greater than 10 and an entropy greater than 6. This is something worthwhile because our DFT-approach not only separates each group from Bach, but it does so for each band together. This means that it understands that Heavy Metal is a style in itself, while Baroque, at least for Bach, is another style. We can assumes from this example that this approach is able to capture two very different styles of composition.

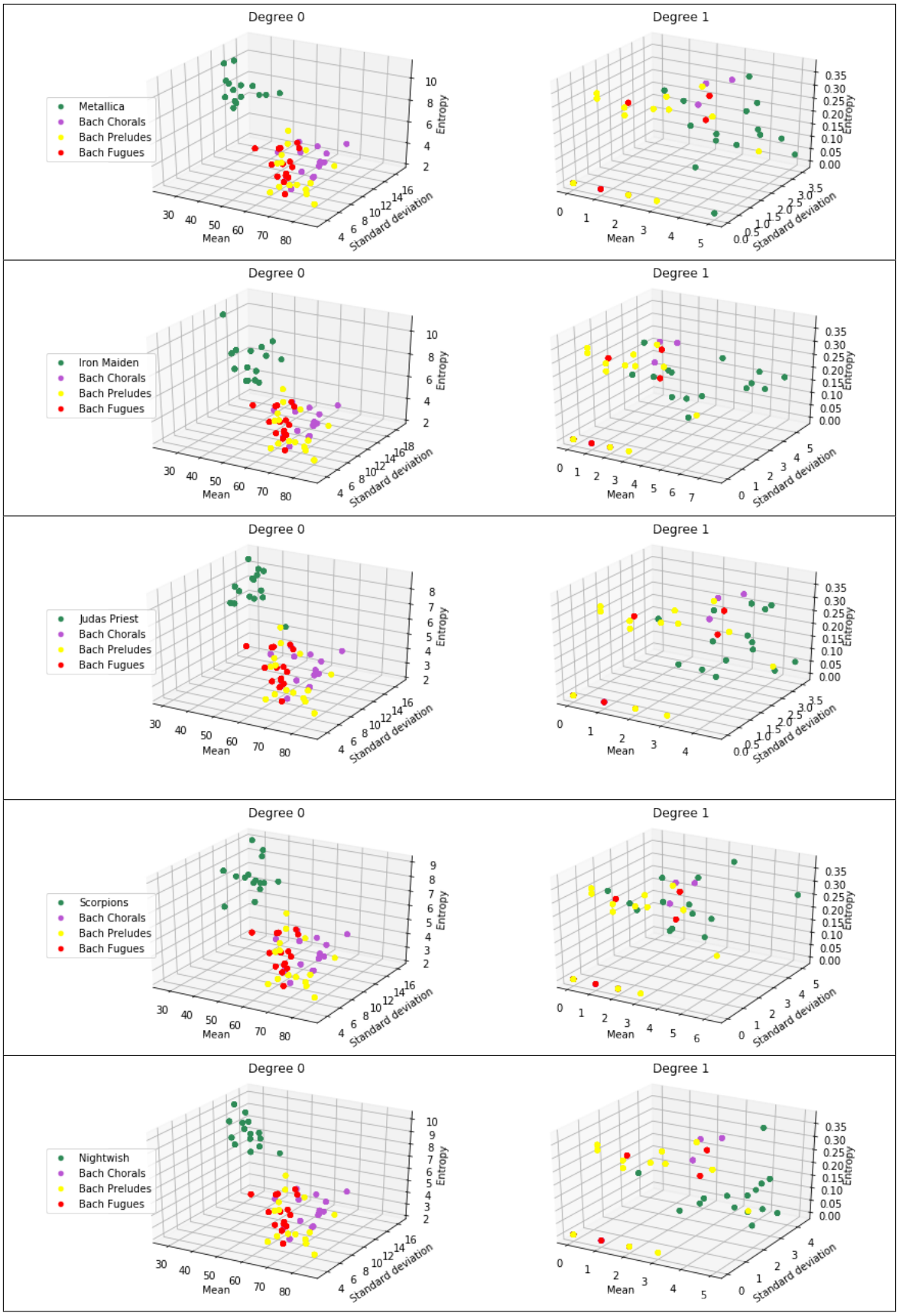


Figure 7.2.1: Metallica, Iron Maiden, Judas Priest, Scorpions and Nightwish compared to Bach (Chorals, Preludes and Fugues): here the results are presented in $\mathbb{R}^3$ and both degree 0 and 1. In the first column, there are clusters that separate each band of Heavy Metal from Bach. On the other hand, there are no obvious clusters in degree 1.

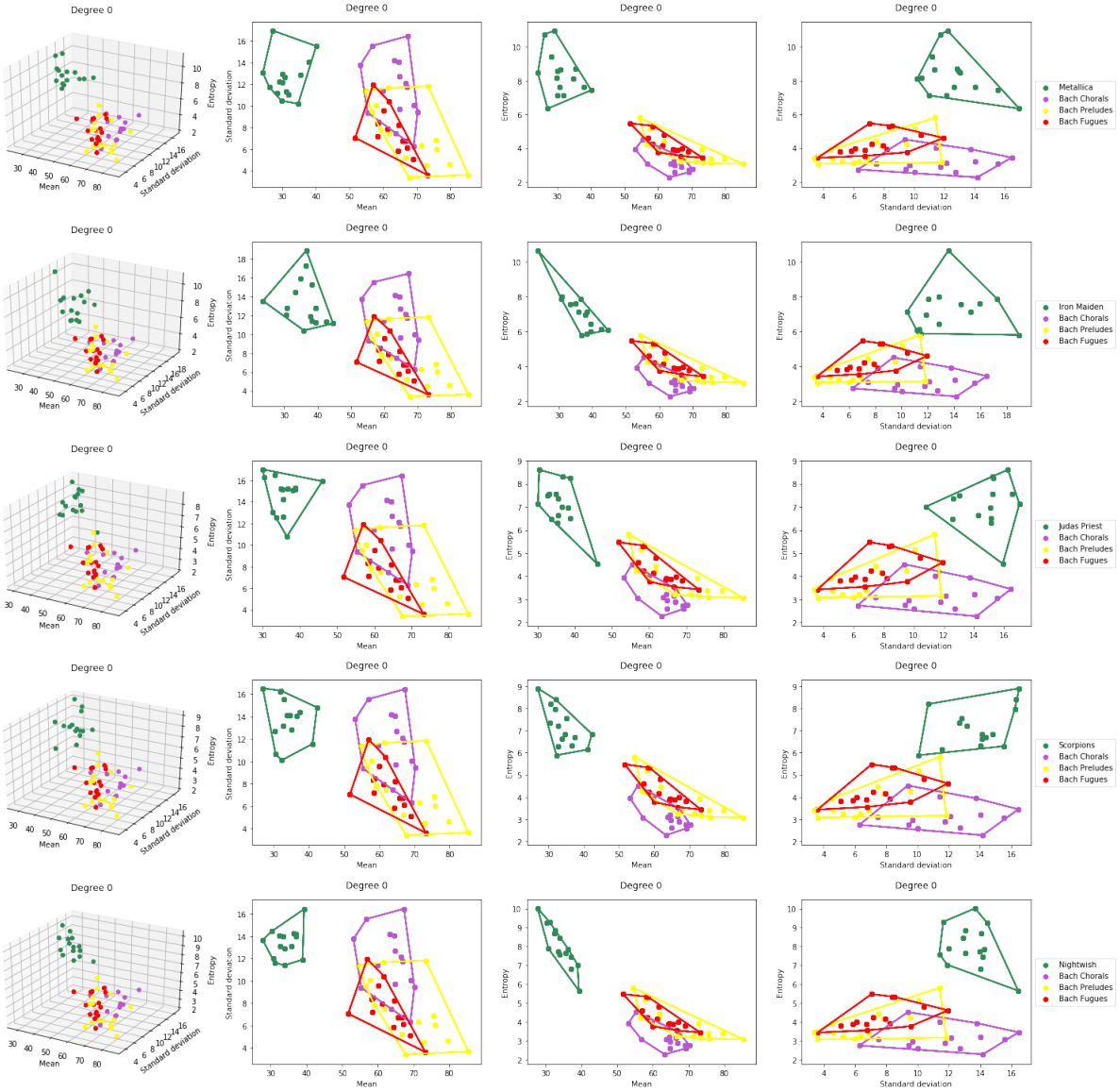


Figure 7.2.2: Metallica, Iron Maiden, Judas Priest, Scorpions and Nightwish compared to Bach (Chorals, Preludes and Fugues) in degree 0 (with projection on each pair of axes): here we show how clusters appear by projecting on each axis and drawing the convex hull around each opponents.

Furthermore, we have seen that the persistent mean and entropy seem to play a particular role in the separation of Heavy Metal and Bach: indeed, it seems that each piece of Metal has a mean of less than 50, while it is the exact opposite for Bach. For the entropy, it is around 5 or 6. A natural question would be to interpret these results in a musical way: what do mean and entropy tell us about the musical style? A small mean corresponds to short bars on average, which seems to be the case for Heavy Metal songs. In degree 0, short bars mean that the different components of the filtration are quickly connected, so these components may look similar and there should be a global structure for the song. On the contrary, for Bach the mean is higher, which corresponds to long bars on average, and it means that components of the filtration remain separated for a long time. This could possibly mean that the DFT-distance captures several parts of the same musical piece. On the other hand, the entropy is a statistical value that measures the average level of surprise or uncertainty that is specific to the studied object. A low entropy means that there is certain amount of redundancy while a high entropy means that there are more surprises and unexpected elements in the piece. For Bach, we have a low entropy, which is related to a possible redundancy of the barcodes while, for Heavy Metal songs, it seems that there is a high level of surprise. It should also be noted that, whereas entropy and mean values have a specific musical meaning, we are not yet able to provide the same kind of interpretation for the standard deviation, which does not seem to exhibit a clear separation between points. However, we can see that all the Heavy Metal songs are clustered in terms of standard deviation, which ranges from 10 to 18, while Bach's pieces have a larger range, from about 4 to 17 (but only until 12 for the Preludes and Fugues).

Finally, we have summarized this first comparison in Figure 7.2.3, which shows how Heavy Metal and Bach form two well-separated clusters at degree 0. It also confirms our previous conclusions: indeed the DFT is able to separate each band of Metal from Bach, which is already satisfactory, but it is also able to say that the different bands of Heavy Metal are similar, since all the corresponding points are mixed together and form a cluster in itself, compared to the one formed by Bach's pieces. Let us see how it works with Classical and Romantic styles.

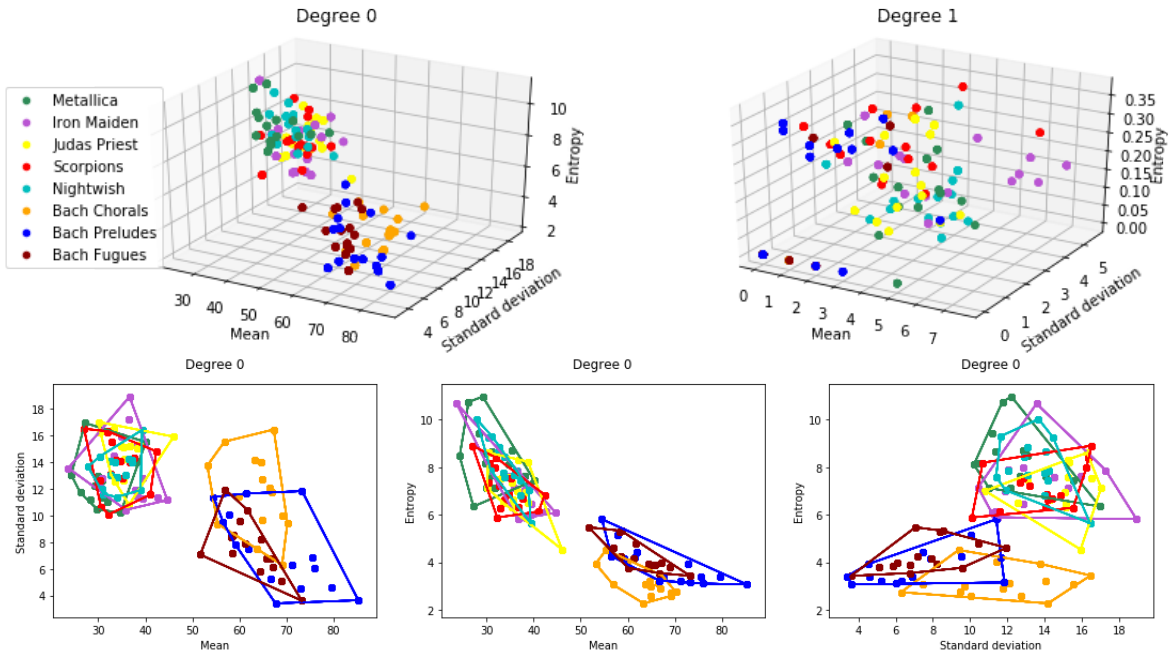


Figure 7.2.3: A synthesis of Heavy Metal compared to Baroque (Bach). At degree 0, the two styles are well separated: the points from Heavy Metal are mixed together to form one cluster, while Chorals, Preludes and Fugues from Bach form another one.

7.2.2. Classical: Wolfgang Amadeus Mozart

The period of Classical musical is roughly between 1750 and 1820, between the Baroque and Romantic periods. This style is known for its sophisticated compositions. During this period, new styles of composition such as Sonatas, Quartet, Concertos and Symphonies appeared. The main instruments used in Classical music are strings (violins, violoncellos, double basses,...), woodwinds (flutes, oboes, clarinets,...) and piano.

As a representative of the Classical style, we will focus on Wolfgang Amadeus Mozart (1756-1791). In the same way we did for Bach, we focus on three different styles of composition from his repertoire: the Piano Sonatas, the Violin Concertos and some vocal works (songs with piano accompaniment). The data used for the analysis are presented in Table 7.2.3 (again, fifteen pieces for each style).

Wolfgang Amadeus Mozart		
Piano Sonatas		
No. 8 1st Mvt, KV311	No. 10 3rd Mvt, KV330	No. 12 2nd Mvt, KV332
No. 8 2nd Mvt, KV311	No. 11 1st Mvt, KV331	No. 12 3rd Mvt, KV332
No. 8 3rd Mvt, KV311	No. 11 2nd Mvt, KV331	No. 13 1st Mvt, KV332
No. 10 1st Mvt, KV330	No. 11 3rd Mvt, KV331	No. 13 2nd Mvt, KV332
No. 10 2nd Mvt, KV330	No. 12 1st Mvt, KV332	No. 13 3rd Mvt, KV332
Violin Concertos		
No. 1 1st Mvt, K207	No. 2 3rd Mvt, K211	No. 4 2nd Mvt, K218
No. 1 2nd Mvt, K207	No. 3 1st Mvt, K216	No. 4 3rd Mvt, K218
No. 1 3rd Mvt, K207	No. 3 2nd Mvt, K216	No. 5 1st Mvt, K219
No. 2 1st Mvt, K211	No. 3 3rd Mvt, K216	No. 5 2nd Mvt, K219
No. 2 2nd Mvt, K211	No. 4 1st Mvt, K218	No. 5 3rd Mvt, K219
Songs		
Daphne, deine Rosenwangen, K52	Die großmüthige Gelaßenheit, KV149	Ich würd auf meinem Pfad, KV 390
An die Freude, K53	Die Zufriedenheit in niedrigen Stande, KV151	Sei du mein Trost, KV 391
O Gotteslamm, K343	Canzonetta Ridenta la calm, KV152	Verdankt sei es dem Glanz, KV 392
Wie unglücklich bin ich nit, KV147	Ariette Oiseaux, si tous les ans, KV307	Ah! spiegarti, oh Dio, KV 417e
O heilges Band, KV148	Ariette Dans un bois solitaire, KV308	(Lied zur) Gesellenreise, KV 468

Table 7.2.3: The forty-five pieces from Piano Sonatas, the Violin Concertos and some vocal works with Songs (with piano accompaniment) from Wolfgang Amadeus Mozart repertory we used for analyzing the Classical style.

Let us focus on degree 0: the comparison between each band of Heavy Metal and the three types of pieces from Mozart is presented in Figure 7.2.4. In this figure, we can immediately see that the separation is less clear than it was for Bach, especially for the last projection. In fact, when the mean value still acts as a separation between the two types of points (around 40), it no longer works well with the entropy. However, we have an interesting fact that comes here: it seems that we have a separation between the Mozart's Songs on the one hand, and his Sonatas and Concertos on the other. In other words, there is a separation between Mozart's compositions per se. This is quite visible on the (mean, entropy)-axis, where the Songs are living in an area delimited by a mean greater than 50 and an entropy less than 5, while Sonatas and Concertos are between 40 and 60 for the mean and 5 and 8 for the entropy. This is musically consistent, since the Mozart's songs we analyzed here are more similar in terms of composition to Bach's Chorals than to his own Sonatas and Concertos, and the DFT-approach seems to capture this difference.

Furthermore, we propose a general comparison in Figure 7.2.5, on which we can confirm that we have clusters given by the mean value (the first two projections). On the other hand, if the (entropy, standard deviation)-axis seems to separate Mozart's Songs from the Metal bands, it is much less clear for the Concertos and the Sonatas, whose entropy value is much higher on average. We underline again the differences between Heavy Metal and Mozart with the first two axes, as well as a difference of style composition for Mozart's pieces studied.

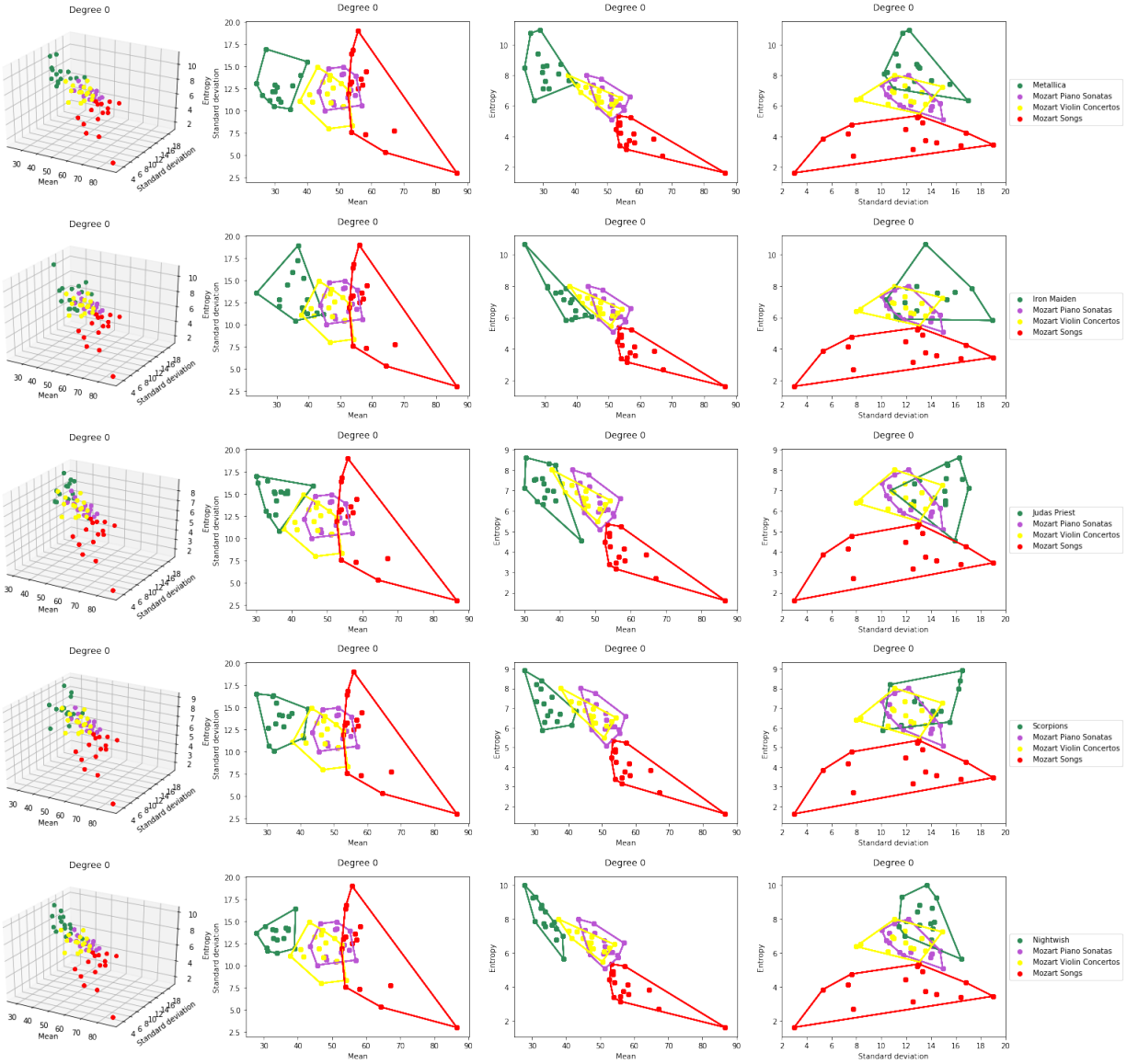


Figure 7.2.4: Metallica, Iron Maiden, Judas Priest, Scorpions and Nightwish compared to Mozart (Piano Sonatas, Violin Concertos, Songs) in degree 0 (with projection on each pair of axes): here we show how clusters appear by projecting on each axis and drawing the convex hull around each opponent. There is also a separation between Mozart's Songs on the one hand, and his Sonatas and Concertos on the other.

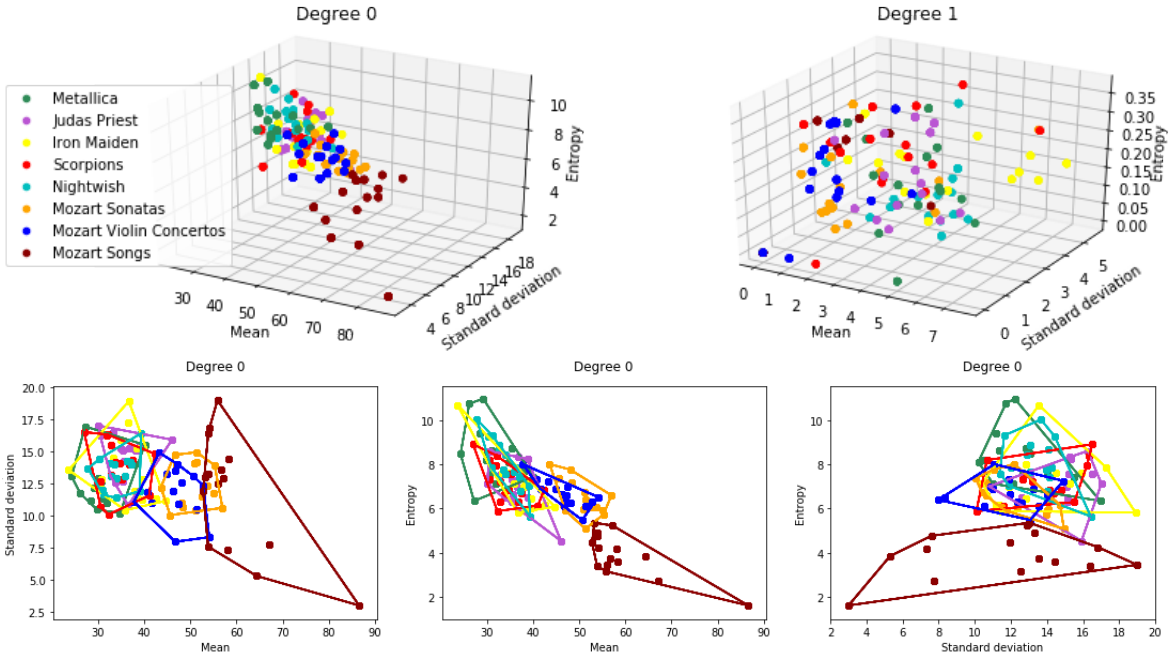


Figure 7.2.5: A synthesis of Heavy Metal compared to Classic (Mozart): in degree 0, there is a separation, especially in the first two projections. There is also a separation between Mozart's Songs on the one hand, and his Sonatas and Concertos on the other.

7.2.3. Romantic: Frédéric Chopin

We conclude this first general analysis with the Romantic style, which is a stylistic movement that comes after the Classical period and lasts during the 19th century, from about 1800 to 1910. Compared to the Classical, Romantic music favours drama and spirituality with much expressive music and also allows for new musical forms such as the rhapsody, the nocturne, the concert étude, the polonaise and the mazurka. The main instruments of this period are piano and the violin.

As a representative of the Romantic style, we have chosen to focus on Frédéric Chopin (1810-1849). More precisely, we focused on three different styles of composition from its repertoire: the Études, the Waltzes and the Nocturnes. The data used for the analysis are presented in Table 7.2.4 (fifteen pieces for each style).

Frédéric Chopin								
Études			**Waltzes**			**Nocturnes**		
Op. 10 No. 1	Op. 10 No. 12	Op. 25 No. 6	Op. 18 No. 1	Op. 64 No. 2	Op. 70 No. 2	Op. 9 No. 1	Op. 15 No. 3	Op. 48 No. 1
Op. 10 No. 3	Op. 25 No. 2	Op. 25 No. 7	Op. 34 No. 1	Op. 64 No. 3	Op. 70 No. 3	Op. 9 No. 2	Op. 27 No. 1	Op. 48 No. 2
Op. 10 No. 4	Op. 25 No. 3	Op. 25 No. 9	Op. 34 No. 2	Op. 69 No. 1	Posth. No. 14	Op. 9 No. 3	Op. 32 No. 1	Op. 55 No. 1
Op. 10 No. 5	Op. 25 No. 4	Op. 25 No. 10	Op. 42 No. 5	Op. 69 No. 2	Posth. No. 15	Op. 15 No. 1	Op. 37 No. 1	Op. 62 No. 1
Op. 10 No. 9	Op. 25 No. 5	Op. 25 No. 11	Op. 64 No. 1	Op. 70 No. 1	Posth. No. 19	Op. 15 No. 2	Op. 37 No. 2	Op. 72 No. 1

Table 7.2.4: The forty-five pieces from *Études*, *Waltzes* and *Nocturnes* from Frédéric Chopin repertory that are used in the analysis of Romantic style against Heavy Metal one.

As we did for Bach and Mozart, we have illustrated the comparison between each Heavy Metal band and Chopin in degree 0 in Figure 7.2.6, and we have also summarized the general comparison in Figure 7.2.7. Here we still find back that most of the Chopin's pieces have a mean greater than 40 (almost 90%), but on the contrary the projection (standard deviation, entropy) seems to completely mix the styles together. However, if we take a closer look at this last graph, we can see that the entropy separates Chopin's Études from all the Heavy Metal

bands, just as it does for Mozart's Songs. It is interesting to note that the Études have a lower entropy than the Nocturnes or Waltzes, and this is musically consistent with the fact that the entropy measures the redundancy of the pieces. Finally, in terms of persistent mean and entropy, Chopin's Études are the pieces that are the closest to Bach's Chorals and Mozart's Songs, while Chopin's Nocturnes show a completely different behavior, which is consistent considering the difference of composition. However, Figures 7.2.6 and 7.2.7 both confirm that we are also able to separate Heavy Metal from Romantic music by focusing especially on the (mean, standard deviation) and the (mean, entropy) projections. These first results lead to a natural comparison between Classical styles per se, which we will discuss in detail in the next section of this chapter.

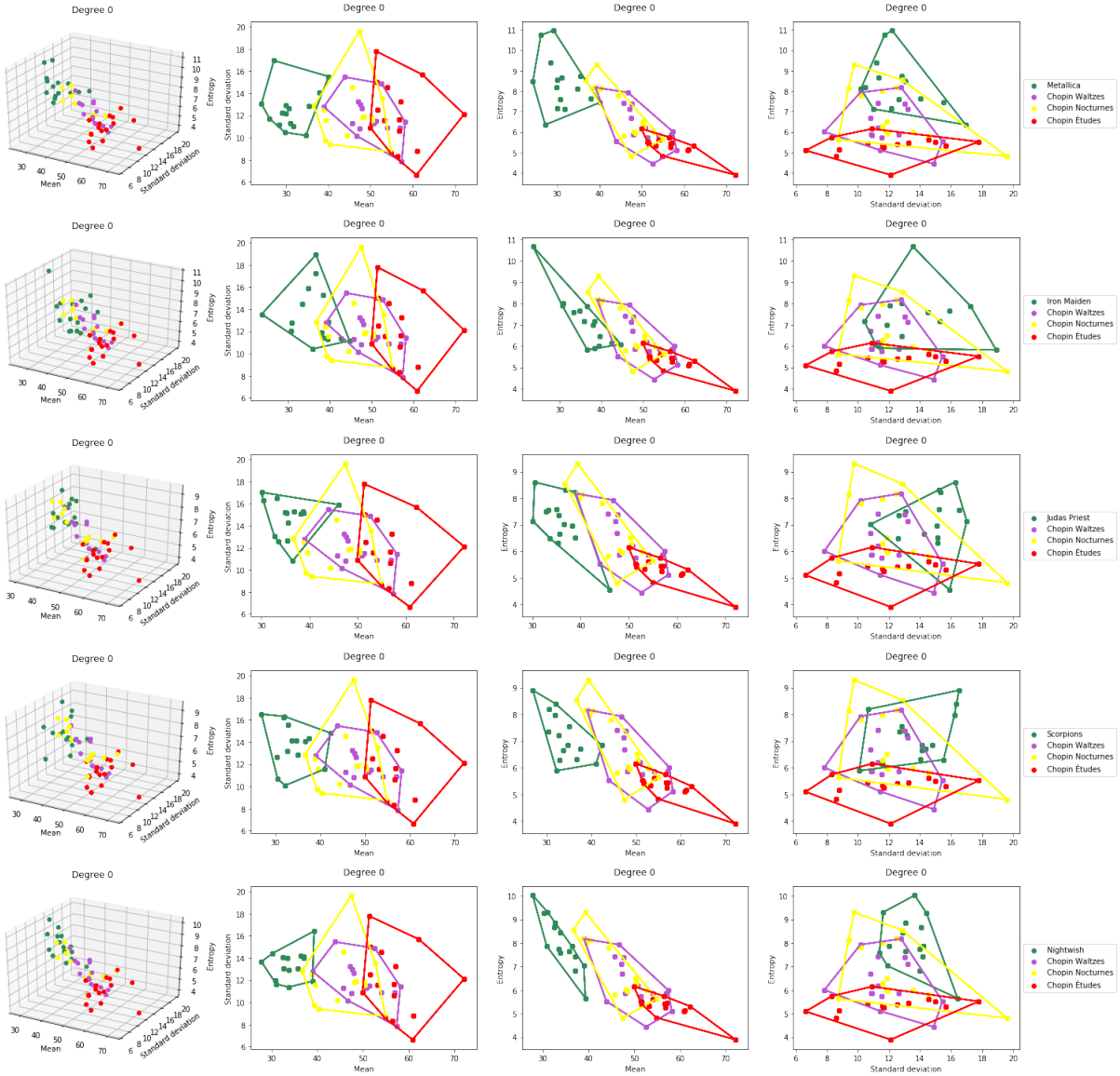


Figure 7.2.6: Metallica, Iron Maiden, Judas Priest, Scorpions and Nightwish compared to Chopin (Études, Waltzes and Nocturnes) in degree 0 (with projection on each pair of axes): here we show how clusters appear by projecting on each axis and drawing the convex hull around each opponent. There is also a separation between Chopin's Études on the one hand, and his Nocturnes and Waltzes on the other.

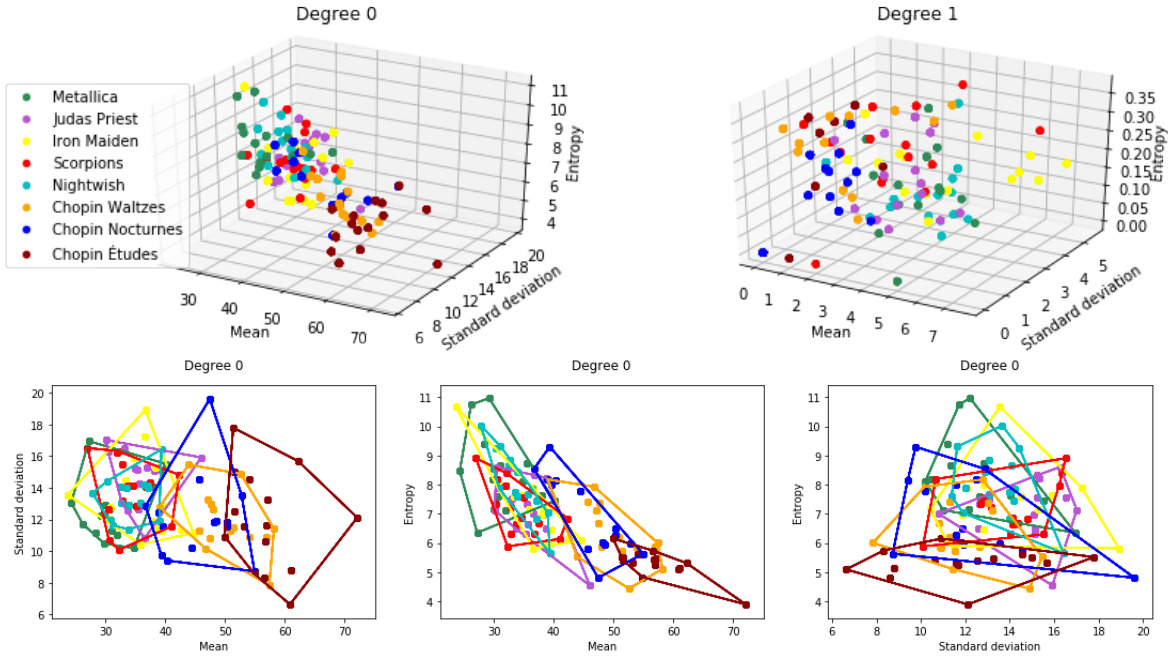


Figure 7.2.7: Heavy Metal compared to Romantic (Chopin): in degree 0, the points of heavy Metal are mixed together and form a cluster, and we have a separation especially given by the mean value. There is also a separation between Chopin's Études on the one hand, and his Nocturnes and Waltzes on the other.

7.3. A study of the Classical style

In the previous section, we have performed first comparison tests between Heavy Metal and Classical style with three sub-genres: Baroque, Classical and Romantic. It turned out that the DFT-approach seems to be able to capture the compositional style of Heavy Metal: in fact, all the songs of each band lived in the same windows, delimited by a mean value less than 50, a standard deviation greater than 10 and an entropy greater than 6. On the contrary, we do not obtain the same results with the Classical pieces studied, as we could observe with Mozart's Songs or Chopin's Études. Therefore, we want to analyze this in depth by looking at the Classical style itself, and this will allow us to refine our approach and see in what the DFT understands from this style. We will also complete this analysis by looking at different styles of composition with Chopin. The data used are the same as in the previous section (Tables 7.2.2, 7.2.3 and 7.2.4).

7.3.1. J.S. Bach, W.A. Mozart and F. Chopin

This paragraph focuses on the Classical style by comparing Bach, Mozart and Chopin together. The results are presented in Figure 7.3.1, where we have performed pairwise comparisons of these styles in degree 0.

The first column shows the comparison between Bach and Mozart and we see that there is indeed a separation between the Sonatas and Concertos from Mozart, which live in a small window delimited by a mean greater than 50, a standard deviation between 8 and 15 and an entropy less than 5. On the contrary, Mozart's Songs seem to completely incorporate the pieces from Bach, with the greatest extent for the standard deviation (from 3 to 20). The separation is present in every projection and seems to place Mozart's Songs in the same compositional style as for Bach's Chorals, Preludes and Fugues.

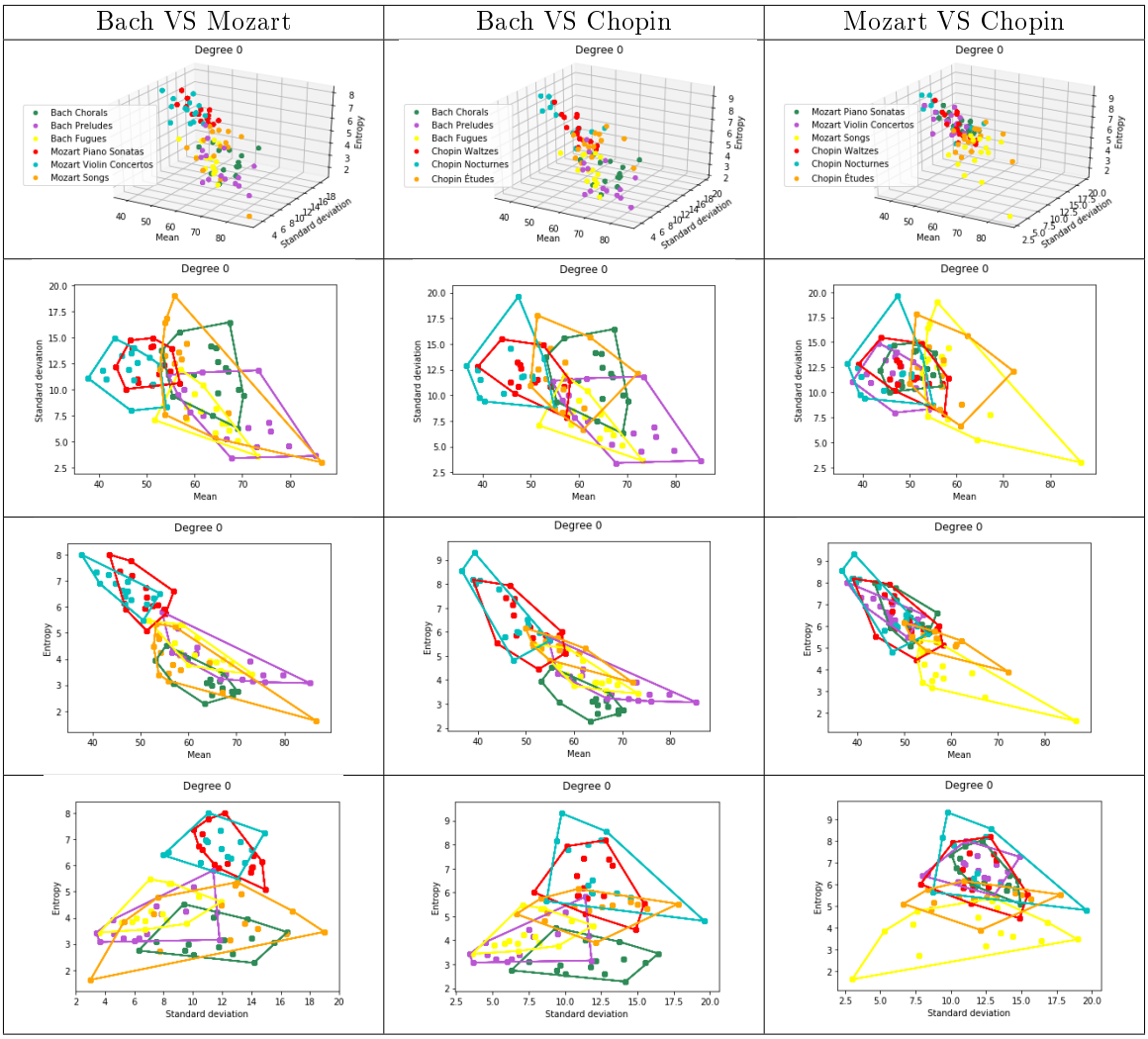


Figure 7.3.1: Pairwise comparisons between Baroque, Classical and Romantic music in degree 0.

We have roughly the same phenomena in the study of Bach against Chopin, and we could observe the separation between Chopin's Nocturnes and Waltzes from Bach. In fact, we still recover a slight separation with a mean of about 50 and an entropy of about 5. On the contrary, Chopin's Études seem to belong to the same compositional style as for Bach's Chorals, Preludes and Fugues, which is also not so surprising since they are much more academic and regulated than the Nocturnes, and therefore the corresponding entropy is lower. Finally, the last column compares Mozart and Chopin, and we still have the idea that here the Classical and Romantic styles are not always separated but the result depends on the studied pieces: here Chopin's Études and Mozart's Songs live in the same windows, delimited by a mean greater than 50, an entropy less then 6 and a large standard deviation. On the other hand, the Piano Sonatas, Violin Concertos, Nocturnes and Waltzes live in the same small window, with a mean between 40 and 60, a standard deviation between 8 and 20 and an entropy between 5 and 9.

In conclusion, the DFT-distance does not separate the studied sub-genres of the Classical style, but it does separate the style of composition, telling us that a Chopin's Étude is closest to a Mozart's Song in terms of compositional style than a Nocturne or a Waltze, which are less codified. In order to support this hypothesis, let us take a final example in the Romantic style, given by Mazurkas from Chopin.

7.3.2. F. Chopin's Mazurkas and Waltzes

The Mazurkas are a Polish musical form based on folk dances in triple time, reminiscent of Waltzes, so it seems to be an interesting approach to confront these two styles We have therefore chosen fifteen Mazurkas from Chopin, which are presented in Table 7.3.1. We keep the original data from Table 7.2.4 and especially the fifteen Waltzes to compare the two styles. We have performed a first comparison between all the Chopin's pieces, and then just between Mazurkas and Waltzes. The results are shown in Figure 7.3.2.

Frédéric Chopin - Mazurkas				
Op. 6 No. 1	Op. 30 No. 4	Op. 41 No. 1	Op. 59 No. 1	Op. 67 No. 1
Op. 7 No. 3	Op. 33 No. 2	Op. 50 No. 1	Op. 59 No. 2	Op. 67 No. 2
Op. 24 No. 4	Op. 33 No. 4	Op. 50 No. 2	Op. 66 No. 3	Op. 68 No. 2

Table 7.3.1: The fifteen pieces from the *Mazurkas* from Frédéric Chopin.

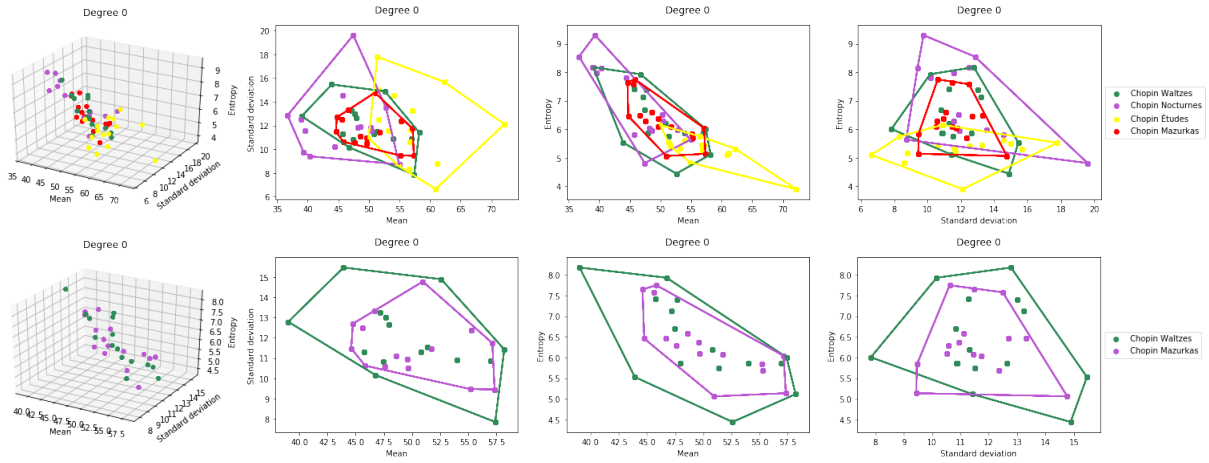


Figure 7.3.2: F. Chopin in four different styles of composition: Nocturnes, Études, Waltzes and Mazurkas. The second line only kept the two last to underline how Mazurkas are included into Waltzes.

In this figure, the first line represents the four compositional styles studied: Nocturnes, Études, Waltzes and Mazurkas. First of all, we can see that the Nocturnes and the Études seem to be different, as it was already mentioned in the previous general analysis. If there is no clear separation, we still have more than 90% of the Nocturnes points having a mean less than 50, while it is greater for 90% of the Études point. We have exactly the same separation with the entropy around 6. This is not so surprising, since among the four types of Chopin's pieces studied here, the Nocturnes and the Études are clearly the most different. This again supports the idea that the DFT is able to capture a style of composition.

The second line of Figure 7.3.2 shows only the Mazurkas against the Waltzes, and there is a surprising fact that appears here: indeed, most of the Mazurkas points are included in the area of the Waltzes points. Therefore, we can assume from this observation that the Mazurkas are a kind of "sub-genre" of the Waltzes in terms of composition, which is a rewarding and musically consistent result that the DFT is thus able to capture.

7.4. What about Pop music?

Now that we have compared trivial opponents and confirmed that the DFT can separate them most of the time, we can focus on less specific genres and turn our attention to Pop music.

Pop music is a genre that originated in the United Kingdom and the United States around 1960. The main idea is to focus on a "popular" accessibility with catchy melodies, dancing rhythms and short songs. By definition, Pop music designates "popular music" so it is something that is constantly evolving over time. The main instruments used in Pop music are electric or acoustic guitars, bass, drums, piano or synthesizers and also drum machines.

There are many choices of groups and artists to represent this style of music, so we decided to cover as wide a time period as possible. We start with The Beatles, who are considered to be the forerunners of pop music in the United Kingdom, and then we move on to nowadays popular music. Here are the five artists we chose to represent pop music:

- **The Beatles** are one of the most famous English band of all the times, formed in Liverpool in 1960, known to be the pioneers of **Pop/Rock style**.
- **Elton John** is a British singer, pianist, composer and representative of **Pop music** of the 80's. His first album was released in 1969 and, he is still in activity.
- **ABBA** are a Swedish group formed in 1972 which are representative of **Pop/Disco style**.
- **Coldplay** are a British band formed in London in 1997, known to represent the **Pop and Pop/Rock style** of the 2000's until nowadays.
- **Lady Gaga** is an American singer and songwriter, who is a representative of **Pop/Dance style**. Her first album was released in 2008, and she is still in activity.

As we did for the Heavy Metal style, we took fifteen songs for each Pop artist, and the data are presented in Table 7.4.1.

At this point, we have Classical music data (Baroque, Classical and Romantic, with forty-five pieces per style), Heavy Metal and now Pop music data (five representatives and fifteen songs for each one). The most natural way to follow this analysis would be to confront Pop music with the two previous styles we have analyzed. For Pop compared to Classical, we expect to find similar results as for Classical compared to Metal. However, the most interesting and non-trivial analysis would probably be the comparison between Pop and Heavy Metal music.

Pop				
The Beatles	**Elton John**	**ABBA**	**Coldplay**	**Lady Gaga**
A Day in the Life (1967)	Can You Feel The Love Tonight ? (1994)	Dancing Queen (1976)	A Sky Full of Stars (2014)	Alejandro (2009)
Across the Universe (1969)	Candle in the Wind (1973)	Does Your Mother Know (1979)	Adventure of a Lifetime (2015)	Bad Romance (2009)
All My Loving (1963)	Crocodile Rock (1972)	Fernando (1976)	Charlie Brown (2011)	Born This Way (2009)
All You Need is Love (1967)	Daniel (1973)	Gimme Gimme Gimme! (1979)	Clocks (2002)	Brown Eyes (2008)
Come Together (1969)	Don't Let The Sun Go Down On Me (1976)	I Have A Dream (1979)	Fix You (2005)	Judas (2011)
Eleanor Rigby (1966)	Goodbye Yellow Brick Road (1993)	Lay All Your Love On Me (1980)	Higher Power (2021)	Just Dance (2009)
Help! (1965)	I Guess That Why They Call It The Blues (1983)	Mamma Mia (1975)	Hymn for the Weekend (2015)	Monster (2009)
Here Comes the Sun (1969)	I'm Still Standing (1983)	Money Money Money (1976)	Paradise (2011)	Paparazzi (2009)
Hey Jude (1968)	Nikita (1985)	People Need Love (1972)	Princess of China (2011)	Poker Face (2008)
Let It Be (1970)	Rocket Man (1972)	SOS (1975)	Sparks (2000)	Rain on Me (2020)
Love You To (1966)	Sacrifice (1989)	Super Trouper (1980)	Speed of Sound (2005)	Shallow (2018)
She Said She Said (1966)	Sad Songs Say So Much (1984)	The Day Before You Came (1982)	The Scientist (2002)	Speechless (2009)
Something (1969)	Sorry Seems To Be The Hardest Word (2002)	The Winner Takes It All (1980)	Trouble (2000)	Summerboy (2008)
Yellow Submarine (1966)	Tiny Dancer (1971)	Voulez-vous (1979)	Violet Hill (2008)	Telephone (2009)
Yesterday (1965)	Your Song (1971)	Waterloo (1974)	Viva La Vida (2008)	The Edge Of Glory (2011)

Table 7.4.1: The five representative Pop bands and the fifteen corresponding songs used in the analysis.

7.4.1. Pop and Classical music

We begin the study of the Pop style by by comparing Pop and Classical music. More precisely, we are going to make the same comparisons that we performed for Heavy Metal, that is, we are going to confront each band of the Pop style successively with Baroque, Classical and Romantic style, by mean of their respective representative that we have chosen in the first section. We will use the data already collected, from Tables 7.2.2, 7.2.3 and 7.2.4 for Bach, Chopin and Mozart respectively, and Table 7.4.1 for Pop music. We have presented the results of each comparison in Figures 7.4.1, 7.4.2 and 7.4.3.

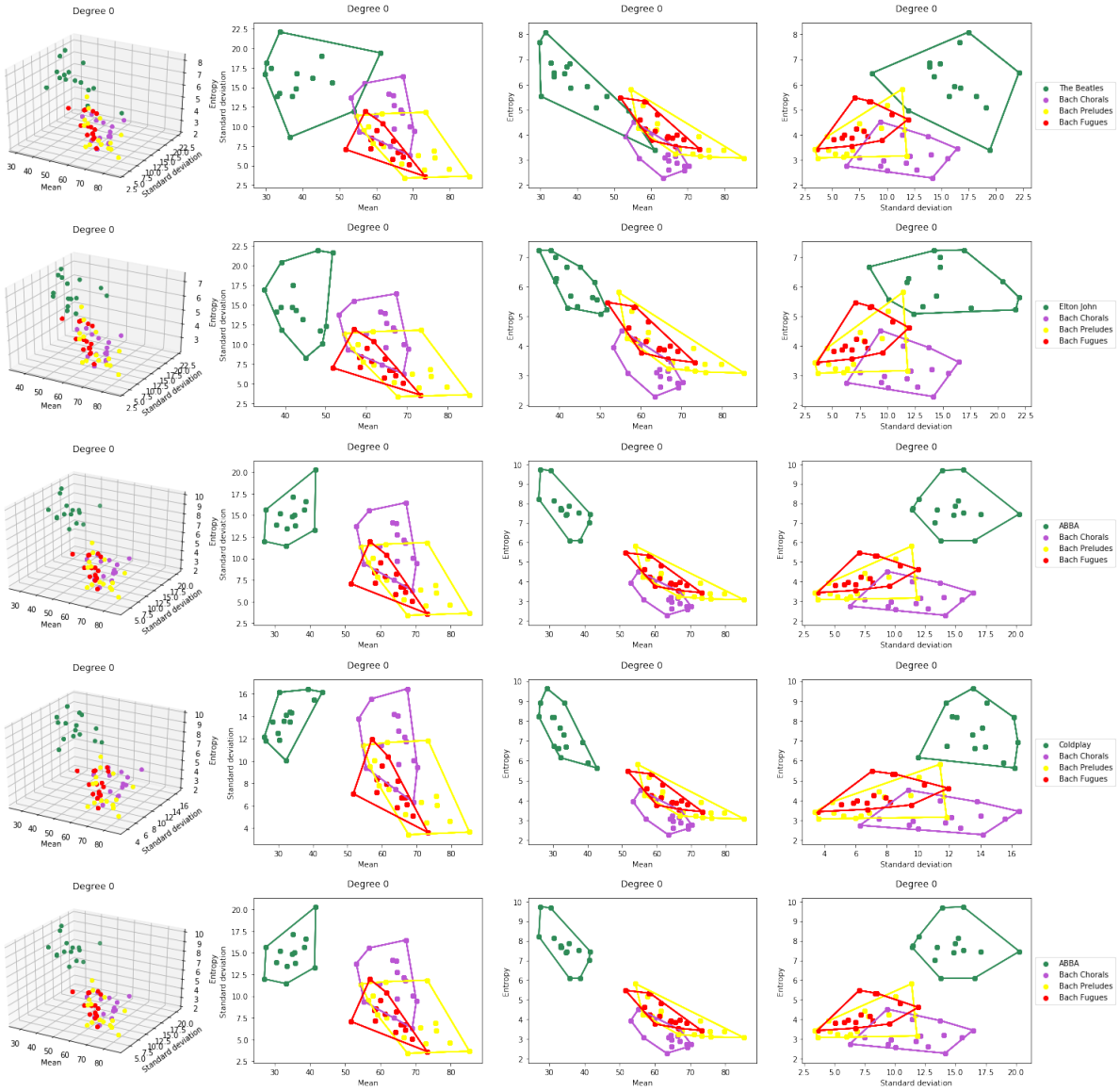


Figure 7.4.1: The Beatles, Elton John, ABBA, Coldplay and Lady Gaga compared to Bach (Chorals, Preludes and Fugues) in degree 0 (with projection on each pair of axes): here we show how clusters appear by projecting on each axis and drawing the convex hull around each opponent. Bach seems to be clearly separated from the Pop bands.

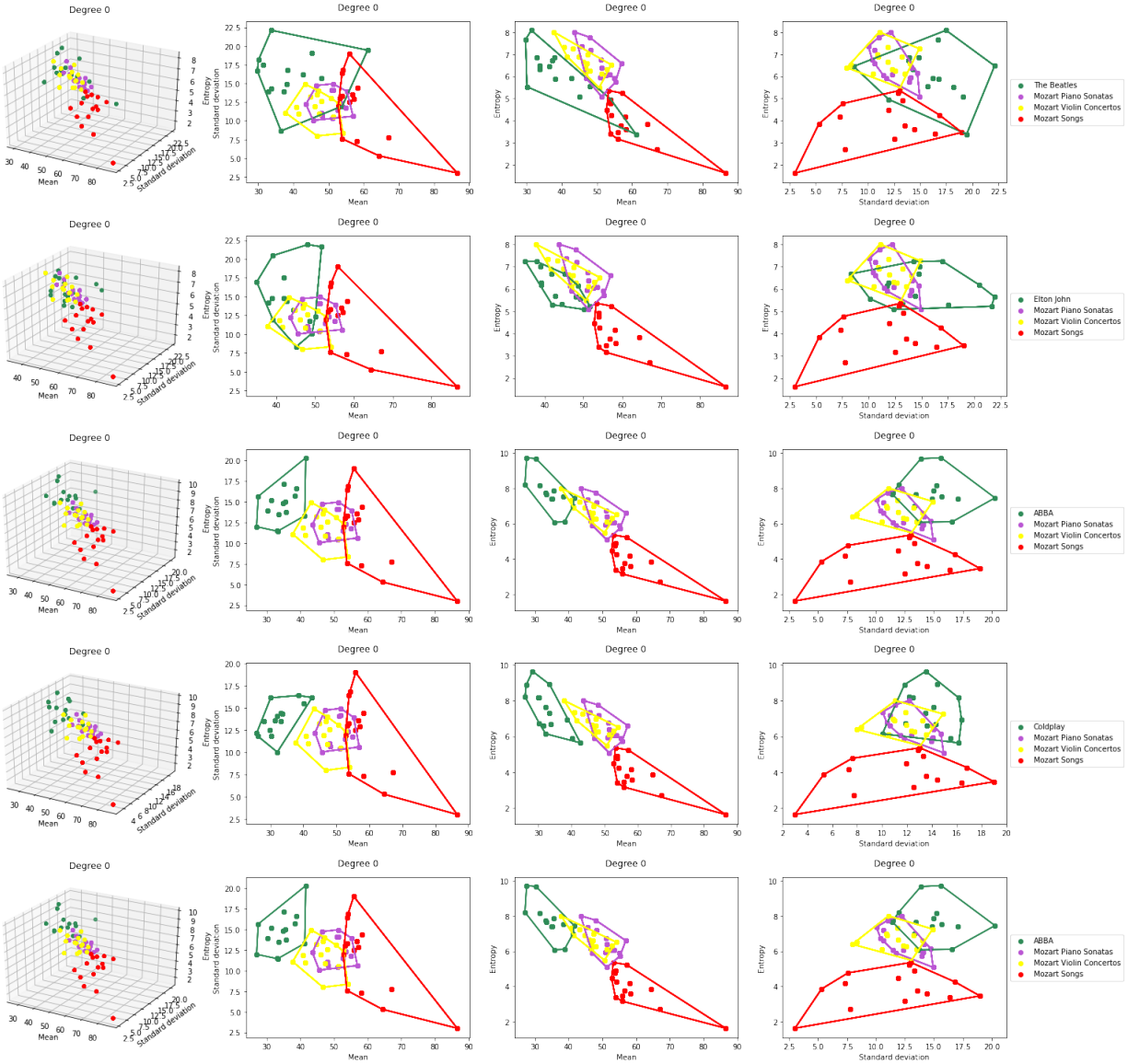


Figure 7.4.2: The Beatles, Elton John, ABBA, Coldplay and Lady Gaga compared to Mozart (Sonatas, Concertos and Songs) in degree 0 (with projection on each pair of axes): here we show how clusters appear by projecting on each axis and drawing the convex hull around each opponent. The first two projections seem to clearly separate Mozart from the Pop bands.

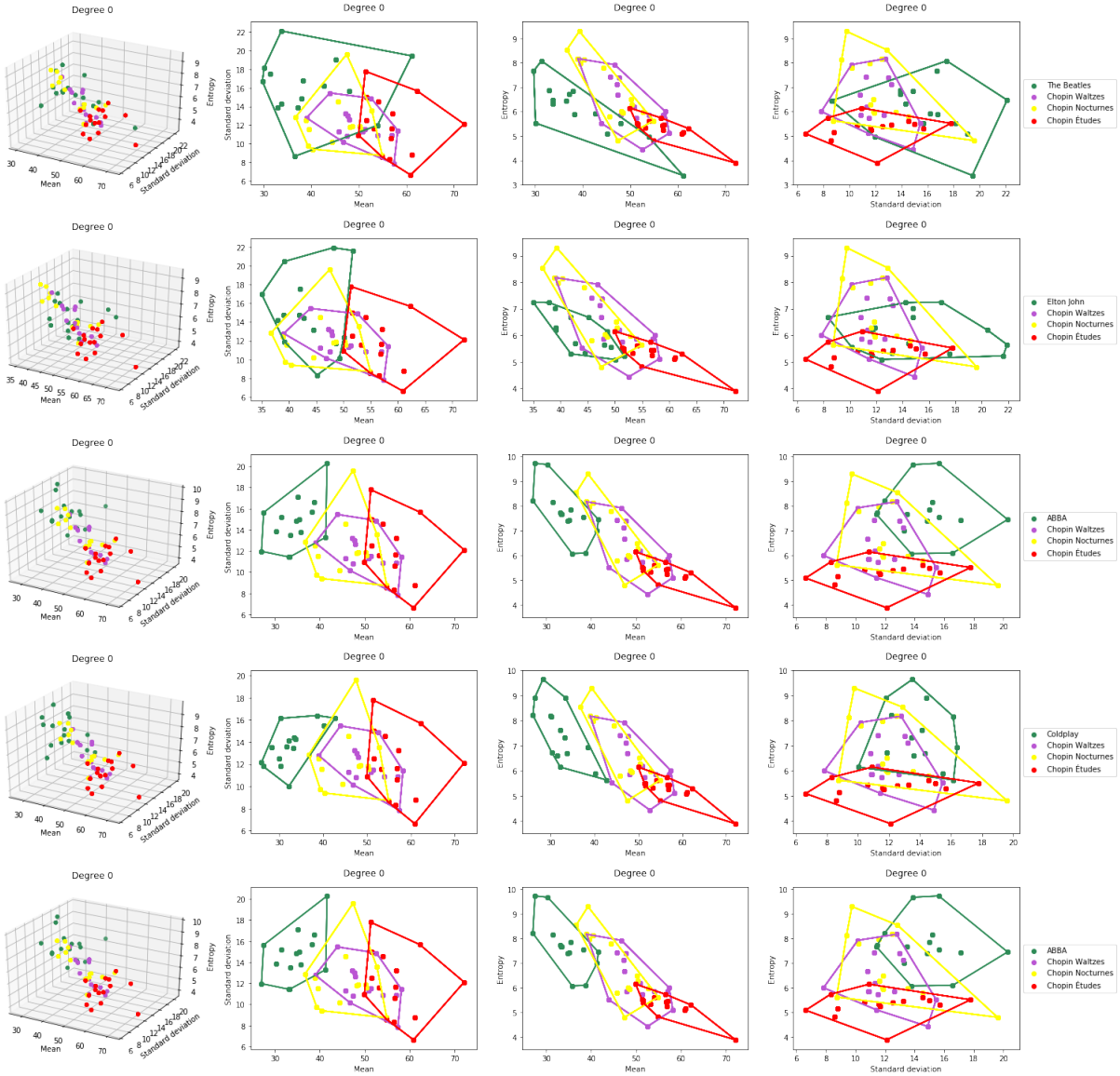


Figure 7.4.3: The Beatles, Elton John, ABBA, Coldplay and Lady Gaga compared to Chopin (Études, Waltzes and Nocturnes) in degree 0 (with projection on each pair of axes): here we show how clusters appear by projecting on each axis and drawing the convex hull around each opponent. The separation is less clear for The Beatles and Elton John, while the first two projections seem to separate the other three Pop bands from Chopin.

As expected, the results are roughly similar to those for Heavy Metal compared to Classical music: for Bach, the separation is clear, especially for ABBA, Coldplay and Lady Gaga, where a mean around 50 and an entropy around 6 seem to create clusters. For Elton John and The Beatles, the mean can go up to 60 while the entropy is very low (around 3 for The Beatles). However, most of the songs studied (more than 95%) are in a different cluster than the pieces of Bach. For Mozart, the projection (mean, entropy) seems to provide a separation between Pop and Classical pieces studied. As it was already the case with the Heavy Metal analysis, the last projection (standard deviation, entropy) seems to mix the points together, especially for the Sonatas and Concertos but still not for Mozart's Songs. It is even more visible in the comparison between Pop and Romantic, where the Nocturnes and also some Waltzes are closed to some Pop songs in the last column. For songs by Elton John and the Beatles, the mean value does not separate the Romantic style as well as it does for Bach and even Mozart. One explanation could that these two artists use the piano mainly as an accompaniment. However, if the clusters are less obvious for Elton John and The Beatles, they are for ABBA, Coldplay and Lady Gaga especially when looking at the first two projections.

7.4.2. A DISTANCE BETWEEN MUSICAL STYLES

The previous section summarizes some comparisons between Pop and Classical music. Most of the time, the results are quite clear and our DFT-distance seems to separate the styles, whether it is with Baroque, Classical or Romantic. However, in some cases it seemed difficult to see clusters in our $\mathbb{R}^3$ representation. In the next section, our aim will be to use the data we have already collected to compare Pop and Heavy Metal together. Therefore, to deal with the problem we encountered with classical music, we first propose a simple and natural way to compare clusters given by the average points.

Since the beginning of this section devoted to the classification issue, we have chosen several musical styles represented by a group or an artist. For each one of them, we took fifteen MIDI files of the corresponding discography, and each song has a barcode that we turned into a point in $\mathbb{R}^3$, using statistical values. Therefore, for two styles, we have two sets of colored points in $\mathbb{R}^3$. In the simplest and most trivial cases, the clusters appear naturally, but sometimes it is not that obvious and we need to be able to perform more complicated comparisons. We thus propose a way to compare groups of points by taking the **average point** in $\mathbb{R}^3$, and more precisely its projection on each pair of statistical axes. We then compare styles by taking the distance between these resulting points. Notice that this distance would also be useful for future work, when we will extend the study and deal with a larger amount of data.

We start with a formalization of the described distance:

Definition 7.4.1. Let $\mathfrak{P}$ be a **family** of musical pieces (or MIDI files), i.e. $\mathfrak{P} = \{\mathcal{S}_1, \dots \mathcal{S}_n\}$ where each $\mathcal{S}_i$ is a score and $n \in \mathbb{N}^*$. For each i, $\mathcal{S}_i$ corresponds to a barcode $\mathrm{BC}_0(\mathcal{S}_i)$ which is identified with a point $X_i = (\mu_i, \sigma_i, \epsilon_i) \in \mathbb{R}^3$, where $\mu_i = \mu_0(\mathcal{S}_i)$, $\sigma_i = \sigma_0(\mathcal{S}_i)$ and $\epsilon_i = \epsilon_0(\mathcal{S}_i)$ are respectively the 0-mean, the 0-standard deviation and the 0-entropy. Therefore, $\mathfrak{P}$ can be associated with a n-uplet $(X_1, \dots, X_n)$, where each X_i is defined as below. We then define the **barycenter of the family** $\mathfrak{P}$ by the average on each coordinate, i.e.

$$\overline{X}_{\mathfrak{P}} = \frac{1}{n}\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i, \sum_{i=1}^{n}\epsilon_i\right).$$

We then denote by $\overline{x}_{\mathfrak{P}}$, $\overline{y}_{\mathfrak{P}}$ and $\overline{z}_{\mathfrak{P}}$ the points in $\mathbb{R}^2$ defined by the projection of the barycenter $\overline{X}_{\mathfrak{P}}$ on each pair of statistical axes, i.e.

$$\overline{x}_{\mathfrak{P}} = \frac{1}{n}\left(\sum_{i=1}^{n}\sigma_i, \sum_{i=1}^{n}\epsilon_i\right), \quad \overline{y}_{\mathfrak{P}} = \frac{1}{n}\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\epsilon_i\right) \quad \text{and} \quad \overline{z}_{\mathfrak{P}} = \frac{1}{n}\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i\right)$$

Now let $\mathfrak{P}$ and $\mathfrak{P}'$ be two families of musical pieces, i.e. $\mathfrak{P} = \{\mathcal{S}_1, \dots \mathcal{S}_n\}$ and $\mathfrak{P}' = \{\mathcal{S}'_1, \dots, \mathcal{S}'_p\}$, where n and p are not necessarily equal. We define the distance between $\mathfrak{P}$ and $\mathfrak{P}'$ by

$$\overline{d}\left(\mathfrak{P}, \mathfrak{P}'\right) = \overline{d}\left(\overline{X}_{\mathfrak{P}}, \overline{X}_{\mathfrak{P}'}\right) = \max\left\{\| \overline{x}_{\mathfrak{P}} - \overline{x}_{\mathfrak{P}'} \|_2, \| \overline{y}_{\mathfrak{P}} - \overline{y}_{\mathfrak{P}'} \|_2, \| \overline{z}_{\mathfrak{P}} - \overline{z}'_{\mathfrak{P}} \|_2\right\}$$

with the notation below and $\| \cdot \|_2$ is the Euclidean norm in $\mathbb{R}^2$. In other words, the distance between $\mathfrak{P}$ and $\mathfrak{P}'$ corresponds to the maximum distance between their respective barycenters on each projection in $\mathbb{R}^2$.

Proposition 7.4.2. *The distance $\overline{d}$ just defined is a metric in $\mathbb{R}^3$.*

Proof. Since our distance $\overline{d}$ is defined with a maximum and the Euclidean norm $\| \cdot \|_2$, we immediately have $\overline{d}(\overline{X}_{\mathfrak{P}}, \overline{X}_{\mathfrak{P}'}) = \overline{d}(\overline{X}_{\mathfrak{P}'}, \overline{X}_{\mathfrak{P}})$ and $\overline{d}(\overline{X}_{\mathfrak{P}}, \overline{X}_{\mathfrak{P}'}) = 0$ if and only if $\overline{X}_{\mathfrak{P}} = \overline{X}_{\mathfrak{P}'}$. Then, using the triangular inequality from $\| \cdot \|_2$, it remains to prove that

$$\max\{a_1 + b_1, a_2 + b_2, a_3 + b_3\} \leq \max\{a_1, a_2, a_3\} + \max\{b_1, b_2, b_3\}.$$

Now assume that the left part of this inequality is given by $a_i + b_i$ for some i, and that the two terms of the right sum are given by a_j and b_k respectively, where i, j and k not necessarily distinct. Then, we have $a_i \leq a_j$ and $b_i \leq b_k$, so $a_i + b_i \leq a_j + b_i \leq a_j + b_k$, and this show the triangular inequality for $\overline{d}$. ∎

Remark 7.4.3. In this definition, we have chosen to work with degree 0 as it seems to be the most consistent dimension to focus on at this stage of our research, but we plan to include higher degrees in our future work. Also notice that we could simply take the distance between the barycenters in $\mathbb{R}^3$ directly, but the previous tests seem to indicate that the choice of the axes in which we project has an influence on the analysis, so we wanted to take this into account in our definition.

The metric we have just defined here is only a suggestion for comparing families of barcodes together, as we decided not to use the Bottleneck distance. This is a research trail that probably has some shortcomings and needs to be exploited, but it has the advantage of using the persistent meaning of barcodes, since it comes discretely from statistical values of bars length. Furthermore, this metric will give some satisfying and musically consistent results (see the clustering trees from the next summary paragraph 7.5).

Example 7.4.4. As a first application of this distance, we have chosen to compare together two bands from Heavy Metal - Scorpions and Nightwish - with two artists from Pop - Lady Gaga and Elton John. In Figure 7.4.4, we show the comparison in $\mathbb{R}^3$ in degree 0 and we add the barycenter point for each group of points (represented by a cross point). Now we can apply our Definition 7.4.1 to compute the distance between the four bands, and the results are presented in Table 7.4.2.

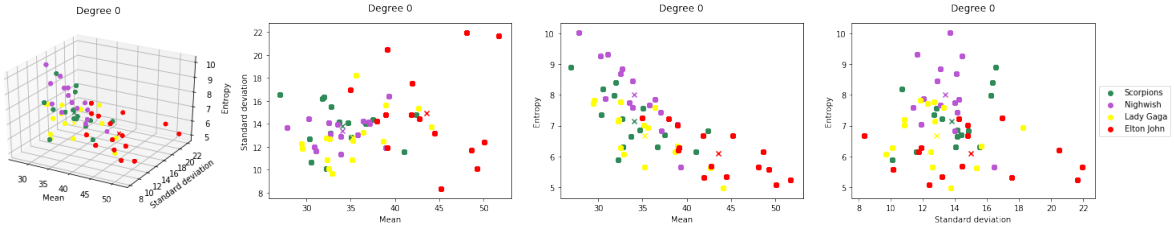


Figure 7.4.4: A comparison between Scorpions, Nightwish, Lady Gaga and Elton John in degree 0. Here we add the average point for each artist.

	Scorpions	Nighwish	Lady Gaga	Elton John
Scorpions	0	0.958	1.542	9.613
Nightwish		0	1.76	9.792
Lady Gaga			0	8.631
Elton John				0

Table 7.4.2: The table of distances between Scorpions, Nightwish, Lady Gaga and Elton John in degree 0 by taking the maximal distances of their average points.

By analyzing Table 7.4.2 of distances, we see that the closest groups are Nightwish and Scorpions, and that Lady Gaga is closer to these two Heavy Metal bands than she is to Elton John. The latter is also the most distant artist from the others, which was already visible in the graphs of Figure 7.4.4, but now confirmed by this new distance.

Finally, we collected the information by drawing the corresponding **clustering tree**, or **dendrogram**, which is shown in Figure 7.4.7. The x-axis corresponds to the computed distances from Table 7.4.4. This representation allows us to immediately visualize the clustering between our studied groups and artists, which is particularly interesting when the classification using visual clusters on $\mathbb{R}^3$ becomes non-trivial or needs to be more precise, as it will be the case with Heavy Metal and Pop music.

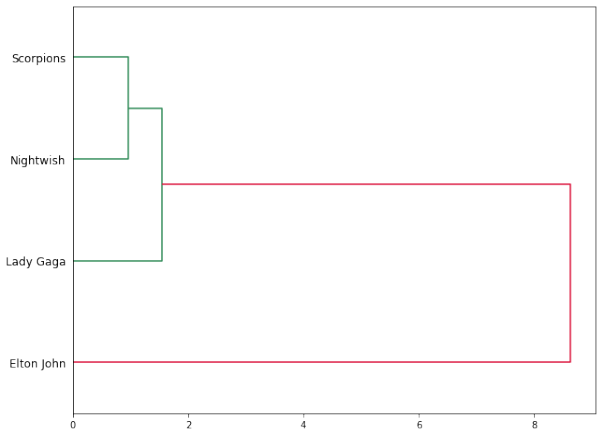


Figure 7.4.5: The dendrogram associated with the comparisons between Iron Maiden, Scorpions, Elton John and Lady Gaga in degree 0.

Remark 7.4.5. Notice that the construction of a dendrogram is equivalent to computing persistent homology in degree 0 and analyzing the evolution of the components during the filtration (as we did in most of our analyses from the previous chapters). In fact, in this context, a point is given by a family of MIDI files represented by its barycenter, according to Definition 7.4.1. The metric between these points is given by the latter definition. From such a point cloud, we can compute the Vietoris-Rips filtration and focus on the grouping of classes in degree 0. This is a "meta-analysis" of the problem, since we compute persistent homology on a family of barcodes.

7.4.3. Pop and Heavy Metal: results and classification

Finally, we computed the different comparisons between each artist or group from our Pop and Heavy Metal data (Table 7.4.1 and 7.2.1 respectively). We give the table of all the different comparisons in the general Figure 7.4.6 for degree 0. For each comparison, we add the the distance between the two groups concerned, as defined in 7.4.1.

Figure 7.4.6: The twenty-five comparisons between each representative of Pop and Heavy Metal music in degree 0. The rows correspond to Metal bands while the columns are for Pop bands.

In these different comparisons we can see that for some groups the clusters are well visible, as for instance in the first two columns (The Beatles and Elton John). On the other hand, for some of the other comparisons, especially the last three columns, it is sometimes not clear. In the previous paragraph, we suggested measuring the distance between groups by taking the average points, projecting them on each pair of axes and taking the maximum difference (see Definition 7.4.1). This allows us to quantify the distance between groups, and also to construct a clustering tree that gives a visual representation of the different clusters. We have thus constructed the corresponding dendrogram with all these comparisons, and the result is presented in Figure 7.4.7. On this dendrogram, we have a cluster between Elton John and The Beatles, who are probably the closest style within the studied Pop groups. Therefore, we have almost a separation between Pop and Metal, except for Lady Gaga, who is between Iron Maiden and Judas Priest, and Metallica, who is between Coldplay and The Beatles.

We conclude this section with a few criticisms of this analysis: first of all, we have to keep in mind that we have completely removed the drums from the MIDI files, which will certainly affect the comparisons, especially when the songs studied comes from Heavy Metal style. Secondly, the clusters necessarily depend on the choice of the songs studied, even though we tried to keep them as diverse as possible: if the fifteen songs by Scorpions are mostly Metal ballads, then without drums it can be really close to songs from a Pop/Rock group like ABBA. This analysis is just the beginning of what can be done by combining persistent homology and the DFT, and it certainly needs to be refined - by adding higher degrees, for example - and tested on a much larger set of data. Nevertheless, we still have consistent results and a search trail in classification

that is promising, especially considering the separation between Pop and Metal that is almost given with this dendrogram. In the next section, we give a list of possible dendrograms that we can construct from our collected data.

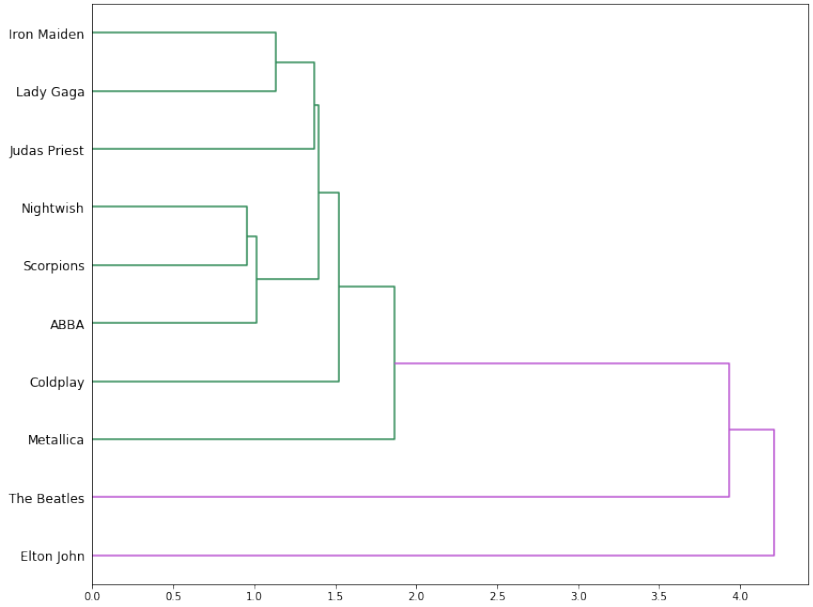


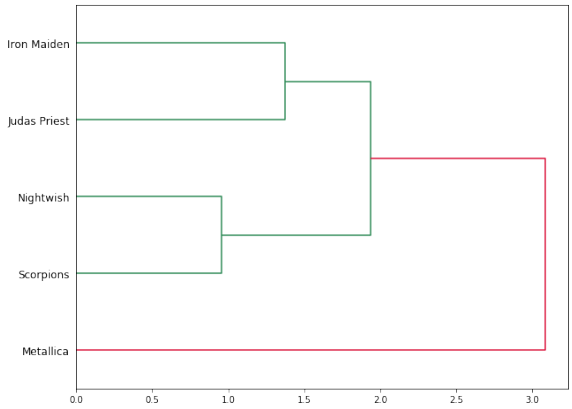
Figure 7.4.7: The dendrogram associated with comparisons between Pop and Heavy Metal music in degree 0. The x-axis is the distance from 7.4.1.

7.5. Synthesis of the various comparisons

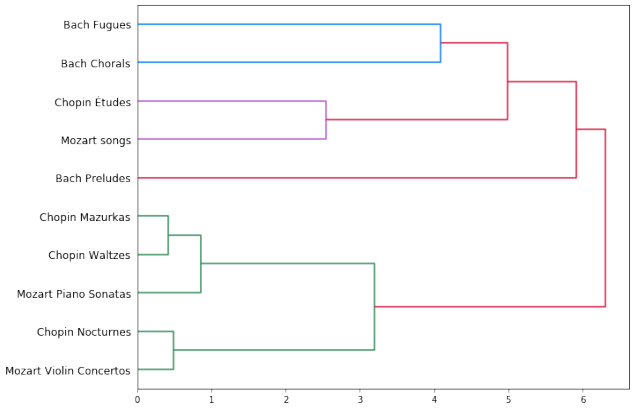
This paragraph simply gives a summary of all the comparisons we can make with the data we have collected by illustrating them with clustering trees. The general results are shown in Figure 7.5.1. We start by giving the dendrogram associated with each style (Figures 7.5.1a, 7.5.1b and 7.5.1c for Metal, Classical and Pop respectively). This allows us to cluster the groups and artists within a given genre. For example, with the Classical one, we recover the cluster given by Chopin's Mazurkas and Waltzes, the one given by Bach's Fugues and Chorals, or also Mozart's Songs and Bach's Chopin's Études. These clusters were all visually given by the $\mathbb{R}^3$ representation and are now confirmed by our new metric. For the Pop style, we recover the clusters given by The Beatles and Elton John.

We also show the clustering tree associated with the Heavy Metal and Classical comparisons from the first paragraph of this chapter (Figure 7.5.1d), and we recover two separate clusters given by these two styles. The comparison between Pop and Classical gives exactly the same results (Figure 7.5.1e). Pop and Heavy Metal have already been compared in the previous paragraph (Figure 7.4.7). Finally, the meta-clustering tree summarized all the possible comparisons between all the 285 studied pieces (Figure 7.5.1f), on which we can clearly see a separation between Classical and the two other styles, and the mix between Pop and Metal. Even if this could be reconstructed from the previous classifications, it also shows that we can immediately apply our approach to a larger data set, and find clusters easily on this representation (compared to the first visualization in $\mathbb{R}^3$), which is promising for future work.

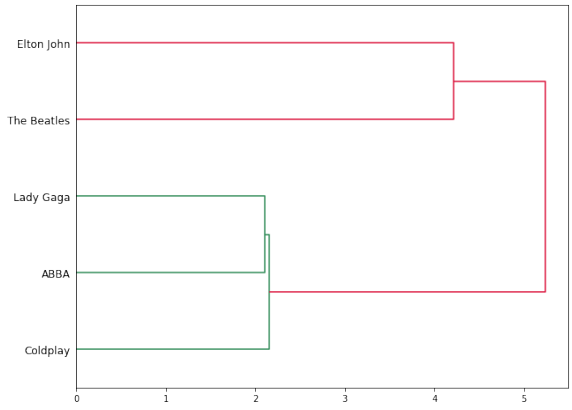
In conclusion, our approach seems to have a real consistent musical meaning, especially in terms of compositional style. The distance from Definition 7.4.1 between families of musical pieces based on their barcodes in degree 0 allows us to compute large comparisons and give a strong visual representation of the works.



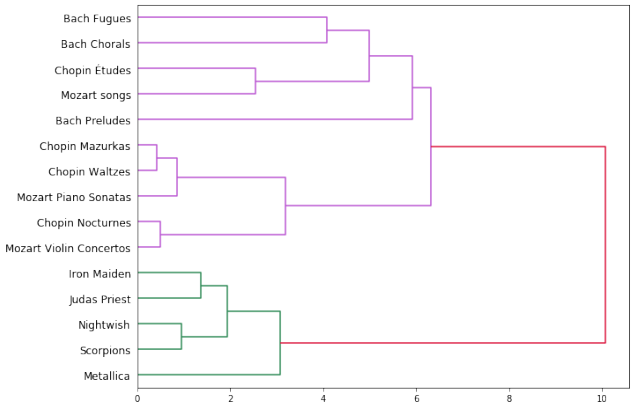
(a) Metal style



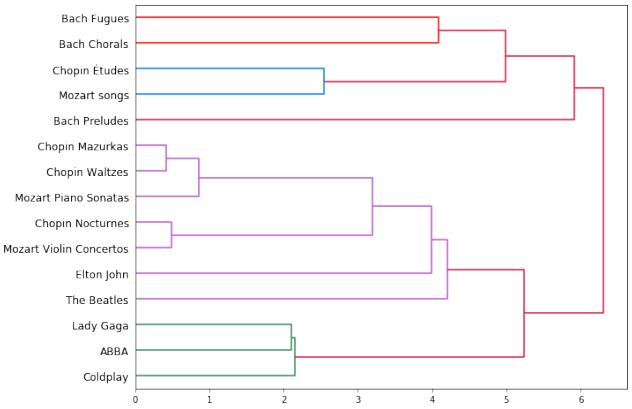
(b) Classical style



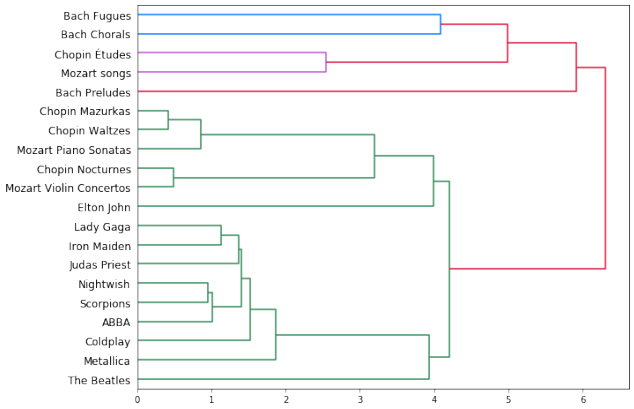
(c) Pop style



(d) Metal VS Classical style



(e) Pop VS Classical style



(f) Classical, Metal and Pop styles

Figure 7.5.1: A synthesis of all comparisons of the different classifications within and between each studied style. The x-axis is the distance from 7.4.1.

7.6. Comparison within one given group: Queen

In this section we propose a different point of view: if our approach can naturally be used to classify several groups or musical styles together, it might be interesting to see what it brings to the study of one given group. For this purpose, we choose to analyze a group with a large discography in terms of musical style: the famous British band Queen.

7.6.1. Different styles for each Queen's member

Queen is one of the most famous British Rock bands, formed in London in 1970. The group is particularly known for its heterogeneous discography, with many different styles throughout its fifteen albums, mainly revolving around Rock. In fact, it goes from Progressive Rock, Hard Rock, sometimes Heavy Metal but mostly Arena Rock and Pop Rock.

The British band played for almost twenty years (from 1973 to 1991) with four members: Freddie Mercury (lead vocals and piano), Brian May (guitar and vocals), Roger Taylor (drums and vocals) and John Deacon (bass). During this time, the group released fourteen albums, and produced a last one in 1995, following Mercury's death in 1991.

Each member of the band has composed for these fifteen albums, and our aim here is to see how our DFT-approach can capture the compositional style that characterized the different pieces. In that purpose, we have chosen fifty-five songs from Queen's discography: twenty of which were composed by F. Mercury, fifteen by B. May, fifteen by R. Taylor and five by J. Deacon. We made this unbalanced selection according to the availability of MIDI files and the proportion of songs composed by each member. The MIDI files we considered are listed in Table 7.6.1.

Queen			
F. Mercury	**B. May**	**R. Taylor**	**J. Deacon**
Bicycle Race (1978)	39' (1975)	A Kind of Magic (1986)	Another One Bites the Dust (1980)
Bohemian Rhapsody (1975)	Fat Bottomed Girls (1978)	Action This Day (1982)	I Want to Break Free (1984)
Crazy Little Things Called Love (1979)	Hammer to Fall (1984)	Calling All Girls (1982)	Need Your Loving Tonight (1980)
Don't Stop Me Now (1978)	Headlong (1991)	Drowse (1976)	Spread Your Wings (1977)
Good Old Fashioned Lover Boy (1975)	I Want it All (1989)	Fun it (1976)	You're My Best Friend (1975)
I'm Going Slightly Mad (1991)	It's Late (1977)	Heaven For Everyone (1995)	
Innuendo (1991)	Keep Yourself Alive (1973)	I'm in Love With My Car (1975)	
It's a Hard Life (1984)	Long Away (1976)	Modern Times Rock 'n' Roll (1973)	
Jealousy (1978)	Now I'm Here (1974)	Radio Ga Ga (1984)	
Killer Queen (1974)	Save Me (1980)	Rock it (1980)	
Liar (1973)	The Show Must Go On (1991)	Tenement Funster (1974)	
Lilly of the Valley (1974)	Tie Your Mother Down (1975)	Thank God It's Christmas (1984)	
Love of My Life (1975)	Too Much Love Will Kill You (1992)	The Invisible Man (1989)	
Made in Heaven (1995)	We Will Rock You (1977)	The Loser in the End (1974)	
March of the Black Queen (1974)	Who Wants To Live Forever (1986)	You Don't Fool Me (1995)	
Play The Game (1980)			
Princes of the Universe (1986)			
Seven Seas of Rhye (1973)			
Somebody to Love (1976)			
We Are the Champions (1977)			

Table 7.6.1: The songs selected from Queen's discography and sorted by composer. There are twenty for F. Mercury, fifteen for B. May, fifteen for R. Taylor and five for J. Deacon.

As a starting point, Figure 7.6.1 shows all the points without any color, simply plotted in $\mathbb{R}^3$ to see how they are distributed. At degree 0, more than 80% of them have a mean between 30 and 50, an entropy between 5 and 9 and a standard deviation between 10 and 16. In degree 1, more than 90% of the points have a mean between 1 and 5 while the standard deviation and the entropy go from 0 to 0.4 for all the points. For the rest of the analysis and as we did for the others before, we will focus only on degree 0.

In Figure 7.6.2, we have represented in the first row the fifty-five points by adding a color for each composer. Considering the low number of points for J. Deacon, we will concentrate mainly

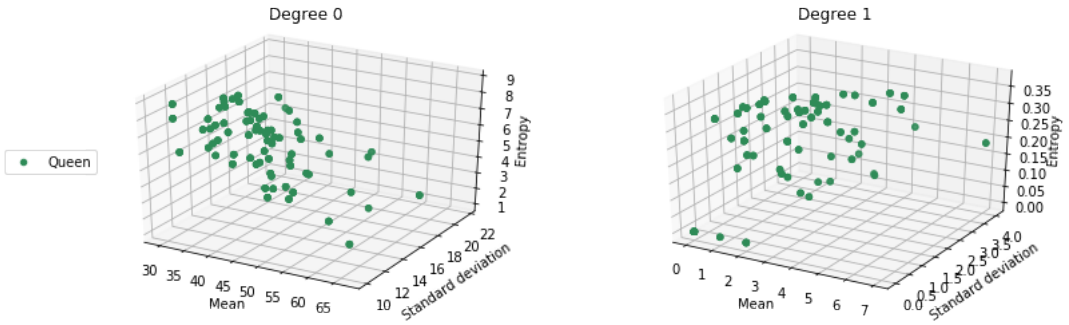


Figure 7.6.1: The fifty-five songs from the discography of Queen in degree 0 and degree 1.

on the other three colors. We can see that most of the points for R. Taylor (70%) have a mean less than 40 and most of the points for F. Mercury and B. May (70%) have a mean greater than 40. For the entropy, we have the opposite: most of the points from B. May have an entropy less than 5.5, while most of the points for F. Mercury and R. Taylor have an entropy greater than 5.5. Therefore, F. Mercury seems to be in the middle of the two styles. We thus decided to display a second line in Figure 7.6.2 where only B. May and R. Taylor are compared.

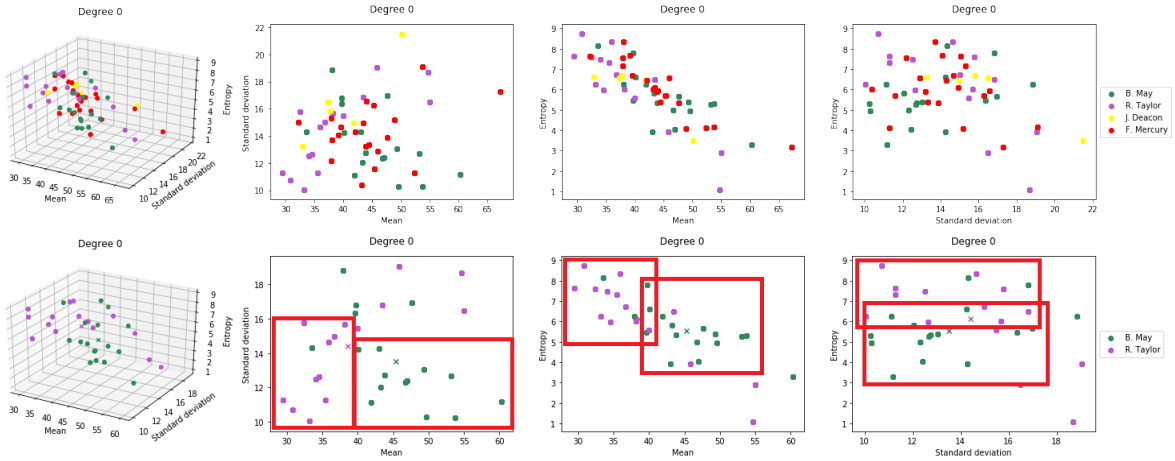


Figure 7.6.2: The fifty-five songs from Queen's discography sorted by composer: in the second line, we have removed F. Mercury and J. Deacon to leave only B. May and R. Taylor. In this case, most of the points are slightly separated, and we plot the region where most of the points lived for each composer.

On these new graphs, we see a slight separation between the two composers, which is given by the mean in the first two projections: the bars of R. Taylor's songs are shortest on average, which probably means that the musical bars are connected quickly and that there is a global structure in the score. On the other hand, most of the B. May's songs have a mean value greater than 40, so the bars are longer on average, and thus there are several pieces in the score that remain separated longer in the filtration. We also have a separation given by the entropy around 5.5, which means that the redundancy in the composition is lower for B. May than it is for R. Taylor. We have displayed these regions for each member and each composer to emphasize the distinction made by the DFT. However, we did not expect the scores to be clearly separated, since B. May and R. Taylor both composed songs for the same group, but the results still show that the DFT distinguishes two styles of composition. Also notice that, in the second row of Figure 7.6.2, we have plotted the barycenter point from Definition 7.4.1 for each color, so we confirm that the points for R. Taylor have on average a mean of less than 40, an entropy of more than 5.5 and also a standard deviation of less than 14, while it is the exact opposite for B. May.

We finally conclude that these two members of the same band have two different compositional styles.

Finally, we confronted these previous results by adding four songs extracted from the discography of B. May during his solo career. In fact, Queen's guitarist released two albums in 1992 and 1998, *Back to the Light* and *Another World* respectively, and we found four MIDI files for these albums: *Driven By You*, *Last Horizon* and *Resurrection* from the former and *Another World* from the latter. We then computed barcodes for these songs and got that all of them have a persistent mean greater than 40 (as 80% of Queen's songs written by B. May) and three of them have an entropy less than 6 (as 70% of Queen's songs written by B. May). The graphical results are presented in 7.6.3.

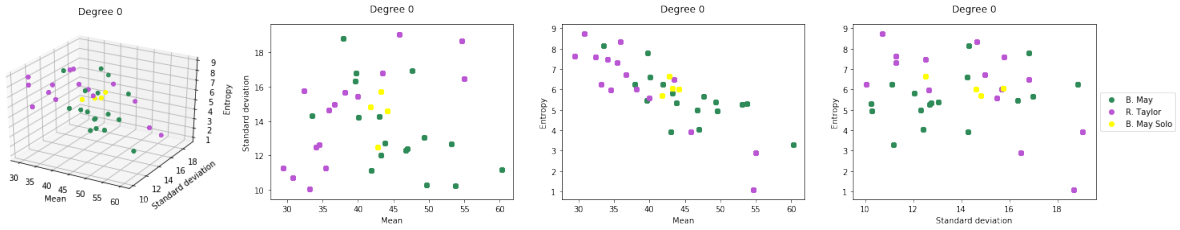


Figure 7.6.3: Thirty songs from the discography of Queen classified by composition (B. May and R. Taylor), together with four songs written by B. May in his solo career.

With this example, we support our belief that DFT can be used in the context of musical style recognition and classification, or more precisely, compositional style, as we have already understood with the previous examples. To conclude this section on Queen, we will again refine our study by applying the DFT to an even more specific case, which is the study of simply a given song from the discography.

7.6.2. Focus on a song: Bohemian Rhapsody

We conclude this section devoted to the classification problem with a paragraph that focuses on just one given song. More precisely, we have made many different comparisons between musical styles, artists and groups and we gave different possible musical interpretations. Here we will try a new approach, by taking just one MIDI file which is itself cut into several parts, in a sense that we will precise. We will then analyze these different parts within the single song.

Since we were working on Queen's discography, we decided to focus on the famous song *Bohemian Rhapsody*, composed by Freddie Mercury in 1975 and released as the lead single from their fourth album *A Night at the Opera* (1975). This song is known to contain several styles in itself: it consists of Rock, Pop and also an Opera part. Our goal is to split this song in order to analyze each part by computing the different statistical values. We will then make a comparison with a pure Thrash Metal song from Metallica, *Master of Puppets*.

Let us start with the cutting of *Bohemian Rhapsody*. The original score contains 154 musical bars (not necessarily distinct), and we split it into six sub-files in the following way:

- **Part 1 :** A Capella introduction, from bars 1 to 16.
- **Part 2 :** Verses 1 and 2 (Ballade), from bars 16 to 48.
- **Part 3 :** Guitar Solo, from bars 48 to 56.
- **Part 4 :** Opera part, from bars 56 to 97.
- **Part 5 :** Rock part, from bars 97 to 137.
- **Part 6 :** Coda ending, from bars 137 to 154.

For each part, we computed different statistical values (in degree 0) and presented the results in the form of histograms, which are shown in Figure 7.6.4. For comparison, the original MIDI file for *Bohemian Rhapsody* has the following statistics in degree 0:

Mean value : 39, **Standard Deviation :** 13.7 and **Entropy :** 7.89.

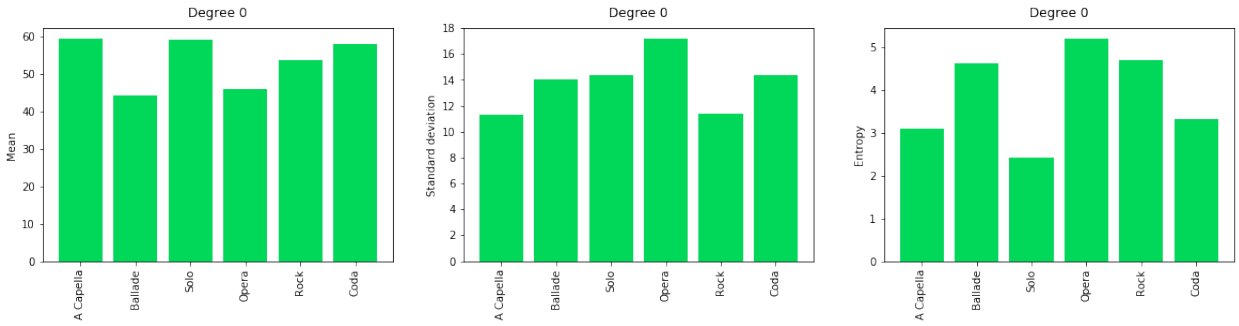


Figure 7.6.4: Mean, standard deviation and entropy in degree 0 for the song *Bohemian Rhapsody*.

Let us analyze these results. First of all, notice that the mean value is always higher than the one for the original song (between 43 and 60), unlike with the entropy which is always lower than 8 (between 2.5 and 5.5). The standard deviation goes from 11 to 17, and the original is in the middle of this interval. What is quite surprising is the way in which a high mean is associated with a low entropy, and vice versa. For instance, the introduction, solo and finale part have the highest mean value (60) but the lowest entropy (around 3), while the opera is one of the smallest mean parts (45) with the highest entropy (5.5). The second thing is the entropy, which is much more variable than the mean or standard deviation: in fact, we notice the difference between the Opera part, which has the highest entropy value at 5.5 and the Solo, which has the lowest at 2.5. In Figure 7.6.5, we give the same histograms in degree 1, which tells us that we do not have many one-dimensional homology generators.

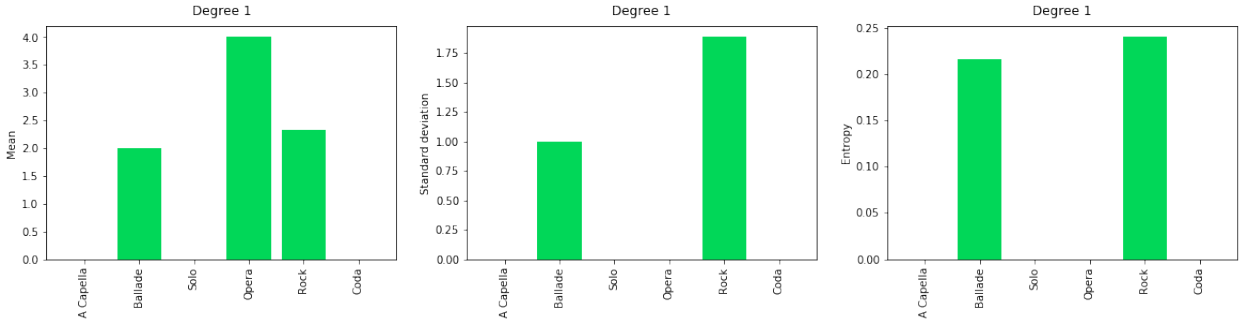


Figure 7.6.5: Mean, standard deviation and entropy in degree 1 for the song *Bohemian Rhapsody*.

Finally, we give a classification of the different parts of *Bohemian Rhapsody* in terms of a clustering tree in Figure 7.6.6, considering the distance from Definition 7.4.1. This classification seems consistent with the statistical values: in fact, it shows that the introduction (A Capella), the ending (Coda), the middle of the song (Solo) and the Pop part (Rock) are successively clustered together, while the Opera and the Pop (Ballade) parts form another group.

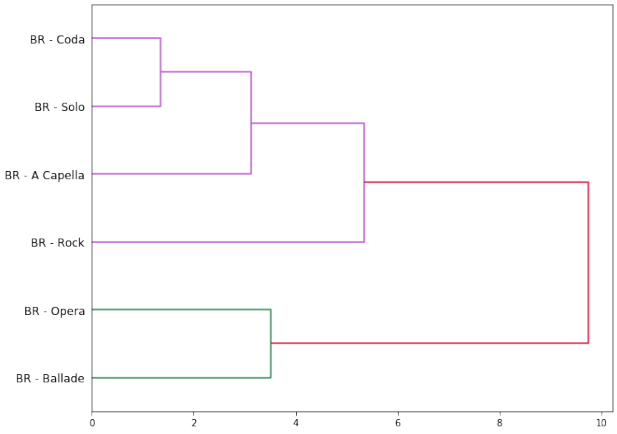


Figure 7.6.6: The clustering tree for the different parts of *Bohemian Rhapsody* (in degree 0). The x-axis is the distance from 7.4.1.

⁕ **A comparison - *Master of Puppets.*** To make a second analysis and compare it with the previous one, we chose to do the same kind of cutting with the Thrash Metal song *Master of Puppets* by Metallica. This song was released in 1986 as the only single from the album of the same name. Notice that it contains much more musical bars than *Bohemian Rhapsody* (401, but not necessarily distinct). We chose this song to have something very different from Bohemian Rhapsody, but still with several parts in it. Therefore, the song can easily be cut in the following way:

- **Part 1 :** Introduction, from bars 1 to 54.
- **Part 2 :** Strophes 1 and 2, from bars 55 to 197.
- **Part 3 :** Ballade, from bars 198 to 238.
- **Part 4 :** Bridge, from bars 239 to 307.
- **Part 5 :** Strophe 3, from bars 308 to 382.
- **Part 6 :** Coda, from bars 383 to 401.

Note that one of the main characteristics of Thrash metal is the use of fast percussive beats, and in our analysis we have suppressed the drums from the MIDI files. Nevertheless, the comparison is probably appropriate, especially considering the clusters formed by Metallica in the previous analysis, which confirm that the group has its particular compositional signature. We give the different statistical values (in degree 0) in Figure 7.6.7. For comparison, the original MIDI file for *Master of Puppets* has the following statistical values in degree 0:

Mean value : 31.1, **Standard Deviation :** 11.33 and **Entropy :** 8.65.

We can immediately obverse that the mean value of this song is more constant than that of Queen, except for the end (Coda), which is even higher (65). The balance with the entropy is also satisfied here: the highest mean value (65) corresponds to the lowest entropy (1.6), given by the Coda. The standard deviation also seems to be more constant (between 12 and 14 except for the Coda which goes to 9). In fact, the analysis reveals that the Coda part is the most different part, which seems musically consistent considering the song. Furthermore, as for *Bohemian Rhapsody*, the entropy for the original song is higher than for the different parts, and vice versa for the mean value, while it seems to be in the middle for the standard deviation. We also illustrate

the same results in degree 1 in Figure 7.6.8, especially to show the difference with *Bohemian Rhapsody* where only 3 parts contain a generator of homology in this dimension. For *Master of Puppets*, almost all of them (except for the Coda) have at least one.

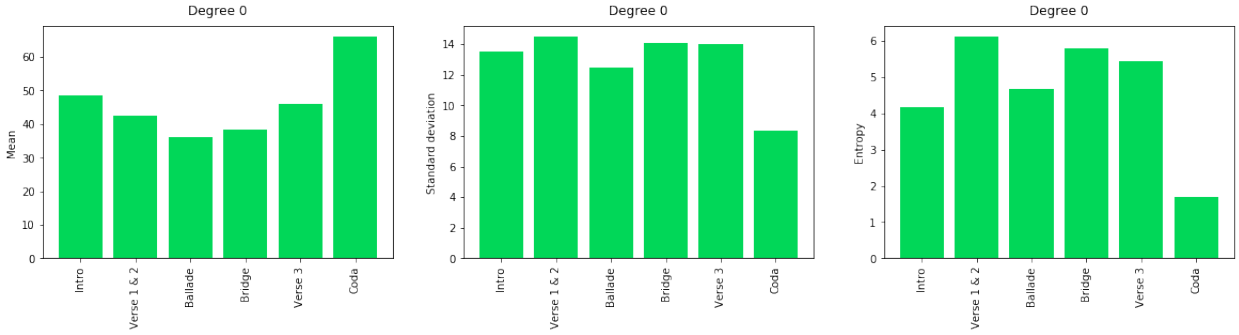


Figure 7.6.7: Mean, standard deviation and entropy in degree 0 for the song *Master of Puppet*.

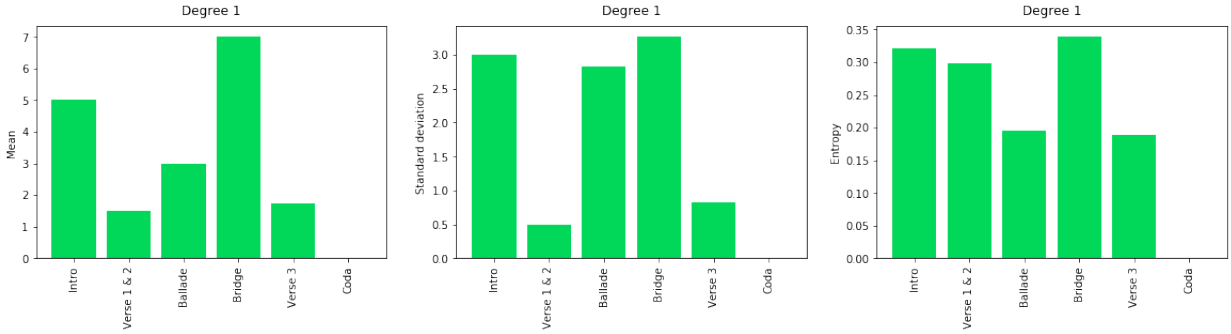


Figure 7.6.8: Mean, standard deviation and entropy in degree 1 for the song *Master of Puppet*.

Finally, we give the corresponding dendrogram in degree 0 in Figure 7.6.9. It is interesting to note that the structure looks quite different than the one from that of *Bohemian Rhapsody*. In fact, we have three main clusters, one given by the verses and the introduction, one given by the bridge and the middle of the song (Ballade), and the Coda, which remains isolated. These clusters are consistent with the statistical values, as it was the case for *Bohemian Rhapsody*.

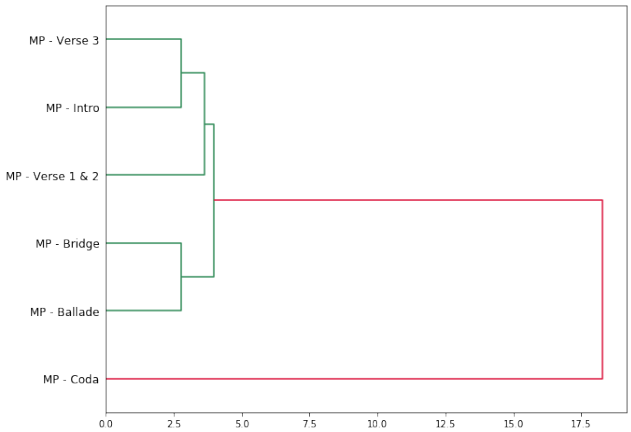


Figure 7.6.9: The clustering tree for the different parts of *Master of Puppets* (in degree 0). The x-axis is the distance from 7.4.1.

To conclude this part, in Figure 7.6.10 we show the corresponding dendrogram in degree 0 with different sections of *Bohemian Rhapsody* together with those of *Master of Puppets*. This clustering tree shows that we have clusters between parts of each song, such as the Coda, Solo, A Capella and Rock parts from *Bohamian Rhapsody*. On the other hand, some parts are mixed together, like the Ballade from the latter, which is placed between the two verses of *Master of Puppets*. However, this new approach based on the DFT and persistent homology seems to be an interesting starting point as a tool which is able to compare two songs in depth.

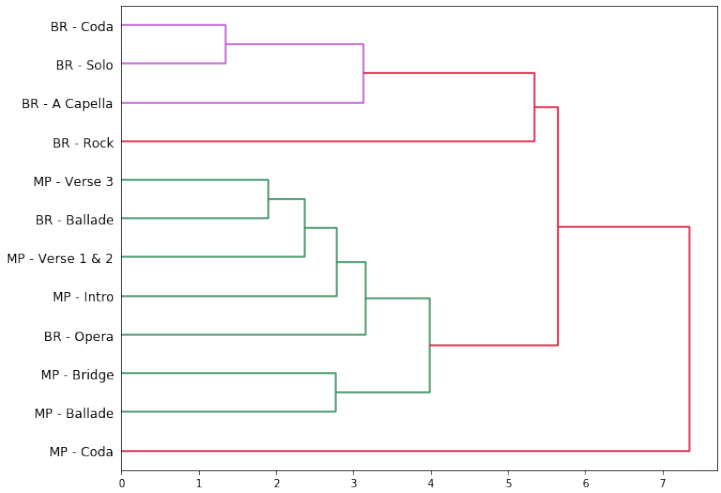


Figure 7.6.10: The clustering tree for the different parts *Bohemian Rhapsody* together with the parts of *Master of Puppets* (in degree 0). The x-axis is the distance from 7.4.1.

Chapter 8.

A different approach: the Hausdorff distance

This chapter is the only part of the manuscript where we will not use the Discrete Fourier Transform to model our scores. More precisely, we keep the idea that a musical piece is represented by a score, which is defined as a finite set of non-ordered distinct musical bars. What is changing here is the modeling of such a bar: in fact, instead of having $\mathcal{B} \subset \mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, we will simply have $\mathcal{B} \subset \mathbb{R}^3$ with coordinates that obviously need to be specified, and we will use the metric given by the Hausdorff distance to compare such musical bars. This model was actually the first we try in our approach of looking at a score as the set of its musical bars. It led to worthwhile results about the global structure of the score and musical texture, shortly presented in [18].

8.1. The model: musical bars in $\mathbb{R}^3$

We keep in mind what we did in Chapter 4: for a musical piece piece $\mathfrak{P}$, let $\mathcal{S}_{\mathfrak{P}}$ be its corresponding score which is cut in N distinct musical bars, so that $\mathcal{S}_{\mathfrak{P}} = \{\mathcal{B}_1, \dots, \mathcal{B}_N\}$. Starting from now, any musical bar $\mathcal{B}_i$ of $\mathcal{S}_{\mathfrak{P}}$ will be living in $\mathbb{R}^3$, according to the following definition.

Definition 8.1.1. A **musical bar** is a finite subset $\mathcal{B}$ of $\mathbb{R}^3$ where an element of $\mathcal{B}$ is called a **note** characterized by three coordinates:

i) the **position** or **onset** which refers to its place in the bar

ii) the **duration**, expressed in beats according to the meter

iii) the **pitch**, which is the value of the note in terms of its fundamental frequency

We are thus considering subsets of $\mathbb{R}^3$ by choosing to represent a note by its onset, its duration and its pitch. This new representation provides a complete and bijective description of any musical bar. One can notice that the first two coordinates are related with **time**, where the third one is about **frequency**. To encode time, we obviously need to use the meter (b, u), where b is the number of beats and u the unit (as we already did in the DFT-case). For instance, if the meter is $(4, 4)$, as it often happens, the unit is the quarter note and there are four such durations in each musical bar of the score (we assume that the musical score has only one meter). As a consequence, a quaver note will by given by $\frac{1}{2}$, a half note by 2 and so on. Also the first two coordinates could never exceed b.

On the other hand, frequencies are encoded using **midicent**, which are basically the unit of frequency in a MIDI file. A small presentation of this encoding is given in Figure 8.1.1 for the three middle octaves of a piano. The middle C, also named C_4, is given by 60 and the midicents go from 0 to 127 (see the introduction part).

We start by giving an example of such an encoding of a musical bar in $\mathbb{R}^3$.

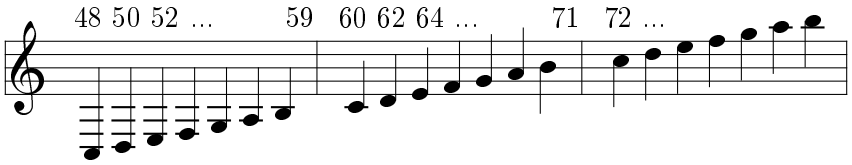


Figure 8.1.1: Encoding the frequency in a MIDI file using the **midicents**. We have C_3, C_4, C_5 and C_6 which are respectively labelled by 48, 60, 72 and 84.

Example 8.1.2. Let us consider the musical bar from Figure 8.1.2.

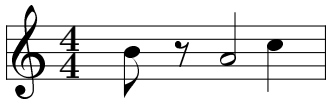


Figure 8.1.2: A musical bar $\mathcal{B}$ encoded in $\mathbb{R}^3$ by $\mathcal{B} = \left\{(0, \frac{1}{2}, 71), (1, 2, 69), (3, 1, 72)\right\}$.

Here the meter is given by $(4, 4)$ so the unit of time is a quarter note and there are four such durations in the bar. More precisely, we have here the following durations:

$$\text{♪} = \frac{1}{2}, \quad \text{♩} = 1 \text{ and } \text{𝅗𝅥} = 2.$$

Concerning frequencies, we have

$$A_4 = 69, \quad B_4 = 71 \text{ and } C_5 = 72.$$

According to Definition 8.1.1, the encoding of this bar $\mathcal{B}$ in $\mathbb{R}^3$ is thus given by:

$$\mathcal{B} = \left\{(0, \frac{1}{2}, 71), (1, 2, 69), (3, 1, 72)\right\}$$

Notice that if there are some rests in the bar, we can ignore them, since the information is already encoded in the position of the next note (which was not the case with the DFT-encoding, where we had no information about the duration of the notes).

Remark 8.1.3. In such encoding, the first two coordinates are really different from the pitch one: in fact, onsets and durations are taking small values depending on the meter, while the frequency encoding with midicent can go from 0 to 127 (most of the time around 60). That was not the case with the DFT-approach where both coordinates of a note were living in $\mathbb{Z}/n\mathbb{Z}$, with an appropriate n. To address this problem, we will make a **normalization** before computing persistent homology, so that we have each musical bar as a subset of $[0, 1]^3$ instead of just $\mathbb{R}^3$.

Remark 8.1.4. As mentioned in Remark 1.2.4 from Section 1.1, there is no bijection between our two-dimensional encoding of a score modulo (t, p) and subsets of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$. On the contrary here, with the Hausdorff approach, there is a complete description of musical bars and scores. Moreover, since we add the duration of a note as a second dimension, we can reconstruct the original musical bar associated with one subset of $\mathbb{R}^3$. Of course, this representation has also some limits, because here we give more weights to coordinates that concern time compared to pitches.

As we did for the DFT-approach and according with our new definition of a musical bar, we can now give the definition of a musical bar in this new context.

Definition 8.1.5. A **musical score** $\mathcal{S}$ is the non-ordered set of its N distinct musical bars in $[0,1]^3$:

$$\mathcal{S} = \{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_N\} \text{ with } \mathcal{B}_i \subset [0,1]^3 \text{ and } \mathcal{B}_i \neq \mathcal{B}_j \text{ if } i \neq j$$

We denote by $\mathfrak{B}^3$ the set of all musical bars in $[0,1]^3$ according to Definition 8.1.1.

Now that we can describe any musical score as a set of finite subsets of $[0,1]^3$, we need to define a metric under the set of these bars that is appropriated. In that purpose, a natural idea is to use a well-known distance to compare metric sets, that means the Hausdorff distance.

Definition 8.1.6. If X and Y are two non-empty subsets of a metric space (M,d), we define a metric d_H called the **Hausdorff distance** between X and Y by:

$$d_H(X,Y) = \max\left\{\sup_{x\in X} d(x,Y) \ ; \ \sup_{y\in Y} d(X,y)\right\}$$

with $d(x,Y) = \inf_{y\in Y} d(x,y)$.

An illustration of this metric is given in Figure 8.1.3.

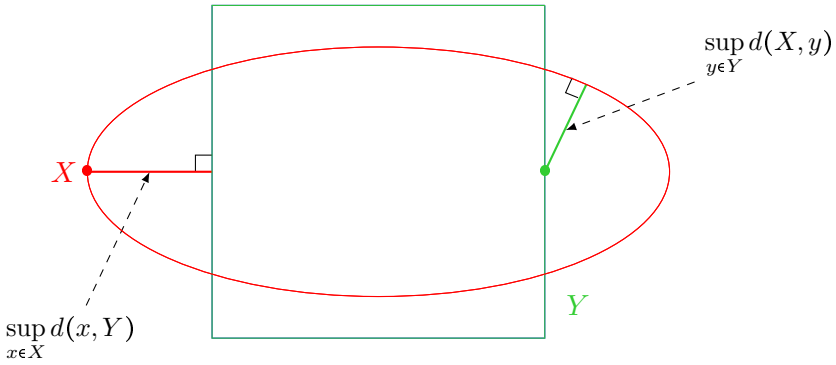


Figure 8.1.3: Calculation of the Hausdorff distance between two metric spaces X (the ellipse) and Y (the square).

This definition leads to a natural metric on the set of musical bars in $\mathbb{R}^3$:

Definition 8.1.7. Let $\mathcal{S} = \{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_N\}$ be a musical score with $\mathcal{B}_i \in [0,1]^3$ for any i. For any pair of musical bar $(\mathcal{B}_i, \mathcal{B}_j)$, their distance is given by

$$d_H^3(\mathcal{B}_i, \mathcal{B}_j) = \max\left\{\max_{n_i\in\mathcal{B}_i} \min_{n_j\in\mathcal{B}_j} d_1(n_i,n_j) \ ; \ \max_{n_j\in\mathcal{B}_j} \min_{n_i\in\mathcal{B}_i} d_1(n_j,n_i)\right\}$$

where $d_1(n_i,n_j) = \sum_i |n_i - n_j|$. We denote by $(\mathfrak{B}^3, d_H^3)$ the metric space of all musical bars encoded as subsets of $[0,1]^3$ equipped with the Hausdorff distance.

Remark 8.1.8. For computational constraints and as we did for the DFT-distance, we chose here to work with the metric on $\mathbb{R}^3$ given by the 1-norm.

We will now turn any musical piece $\mathfrak{P}$ as a point cloud, so we can apply the Vietoris-Rips method from Chapter 4, and thus compute persistent homology and associated family of barcodes.

Definition 8.1.9. Let $\mathfrak{P}$ be a musical piece and $\mathcal{S}_\mathfrak{P} = \{\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_N\}$ the score that represents it. Then, $\mathcal{S}_\mathfrak{P}$ is a subset of the metric space $(\mathfrak{B}^3, d_H^3)$ and the **associated point cloud** with $\mathfrak{P}$ is built in the following way:

i) Each **point** is a musical bar $\mathcal{B}_i \subset [0,1]^3 3$ (Definition 8.1.1).

ii) The **distance** between two musical bars $\mathcal{B}_i$ and $\mathcal{B}_j$ is given by the Hausdorff distance d_H^3 (Definition 8.1.7).

The previous definition leads to the application of Topological Data Analysis in this context, according to the process described in Chapter 4. The next section of this chapter will be devoted to such applications, but first we will give an example of the general process that leads from a musical score to a family of barcodes, according to our new model.

Example 8.1.10. Let us consider the score $\mathcal{S}$ from Figure 8.1.4 which is an excerpt from the music piece *Laputa: Castle in the Sky* composed by Joe Hisaishi in 1986 for the movie of the same name.

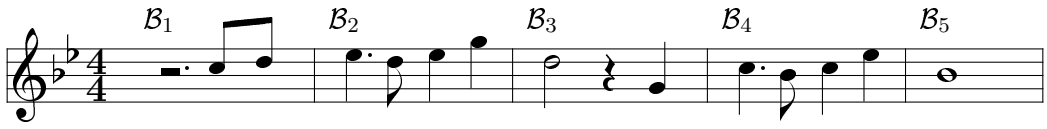


Figure 8.1.4: A score extracted from *Laputa: Castle in the Sky* written by Joe Hisaishi in 1986 for the same called movie.

Following on 8.1.1 and 8.1.5, this score is entirely represented with subsets of $\mathbb{R}^3$ by:

$$\begin{aligned}
\mathcal{B}_1 &= \left\{(3, \tfrac{1}{2}, 72),\ (\tfrac{7}{2}, \tfrac{1}{2}, 74)\right\} \\
\mathcal{B}_2 &= \left\{(0, \tfrac{3}{2}, 75),\ (\tfrac{3}{2}, \tfrac{1}{2}, 74),\ (2, 1, 75),\ (3, 1, 79)\right\} \\
\mathcal{B}_3 &= \{(0, 2, 74),\ (3, 1, 67)\} \\
\mathcal{B}_4 &= \left\{(0, \tfrac{3}{2}, 72),\ (\tfrac{3}{2}, \tfrac{1}{2}, 59),\ (2, 1, 70),\ (3, 1, 75)\right\} \\
\mathcal{B}_5 &= \{(0, 4, 59)\}
\end{aligned}$$

Then, after doing the normalization process for each coordinate (Remark 8.1.3), we get that $\mathcal{B}_i \subset [0,1]^3$ for each i, as follows:

$$\begin{aligned}
\mathcal{B}_1 &= \{(0.75, 0.118, 0.586),\ (0.875, 0.118, 0.609)\} \\
\mathcal{B}_2 &= \{(0, 0.356, 0.621),\ (0.375, 0.118, 0.609),\ (0.5, 0.237, 0.628),\ (0.75, 0.237, 0.667)\} \\
\mathcal{B}_3 &= \{(0, 0.474, 0.609),\ (0.75, 0.237, 0.529)\} \\
\mathcal{B}_4 &= \{(0, 0.356, 0.586),\ (0.375, 0.118, 0.563),\ (0.5, 0.237, 0.586),\ (0.75, 0.237, 0.621)\} \\
\mathcal{B}_5 &= \{(0, 0.949, 0.563)\}
\end{aligned}$$

Now we can simply use the Hausdorff distance from Definition 8.1.7 and Table 8.1.1 of distances (left). In order to compare barcodes together, we apply the same distances normalization process than in the DFT context (see Section 4.3). We then get every distances between 0 and 100 as presented in Table 8.1.1 (right): the time of filtration is now measuring with **scaling parameter** (%).

Finally, we have our point cloud and we can now apply the Topological Data Analyis process, that means computing the corresponding filtration and family of barcodes. The results are presented in Figures 8.1.5 and 8.1.6, respectively.

Let us briefly analyze this example: here we do not have homology in dimension 1, which is not so surprising considering the low number of musical bars in the score. However, there is

	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_3$	$\mathcal{B}_4$	$\mathcal{B}_5$
$\mathcal{B}_1$	0	1.02	1.13	0.987	1.75
$\mathcal{B}_2$		0	0.574	0.045	1.57
$\mathcal{B}_3$			0	0.528	1.5
$\mathcal{B}_4$				0	1.52
$\mathcal{B}_5$					0

	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_3$	$\mathcal{B}_4$	$\mathcal{B}_5$
$\mathcal{B}_1$	0	58%	64%	56%	100%
$\mathcal{B}_2$		0	32%	2%	89%
$\mathcal{B}_3$			0	30%	85%
$\mathcal{B}_4$				0	86%
$\mathcal{B}_5$					0

Table 8.1.1: Tables of distances before and after distances normalization (respectively left and right) for the musical score $\mathcal{S}$ from *Laputa: Castle in the Sky*. On the right we simply apply the DFT normalization process from Section 4.3.

an interesting process that should be noticed here, since it will be the starting point of the next paragraph: the first two musical bars that are connected together are $\mathcal{B}_2$ and $\mathcal{B}_4$, with a very small distance (a scale of 2%). Then, we have $\mathcal{B}_2$, $\mathcal{B}_3$ and $\mathcal{B}_4$ that form a triangle while $\mathcal{B}_1$ and $\mathcal{B}_5$ are still isolated (at 32%). Finally, if $\mathcal{B}_1$ joins the triangle at 56% of the filtration, and we need to wait 85% before having the same property for $\mathcal{B}_5$. In other words, half of the time $\mathcal{B}_1$ and $\mathcal{B}_5$ are two isolated components, and quarter of the time $\mathcal{B}_5$ is a single one. This progression has a very particular musical meaning: firstly, $\mathcal{B}_2$ and $\mathcal{B}_4$ are almost the same up to a transposition, and at least these two musical bars share the exact same **rhythmic structure**. We will say that $\mathcal{B}_2$ is the **theme part** that is reused in $\mathcal{B}_4$. Then, $\mathcal{B}_1$ sounds like the **beginning** of the musical sentence, while $\mathcal{B}_5$ is the **conclusion**, and $\mathcal{B}_3$ acts as a **transition** (or **bridge**). With this small example, we see the premises of the results we can obtain with the Hausdorff distance, that is to say, describe in an automatic way the general **structure** of a piece.

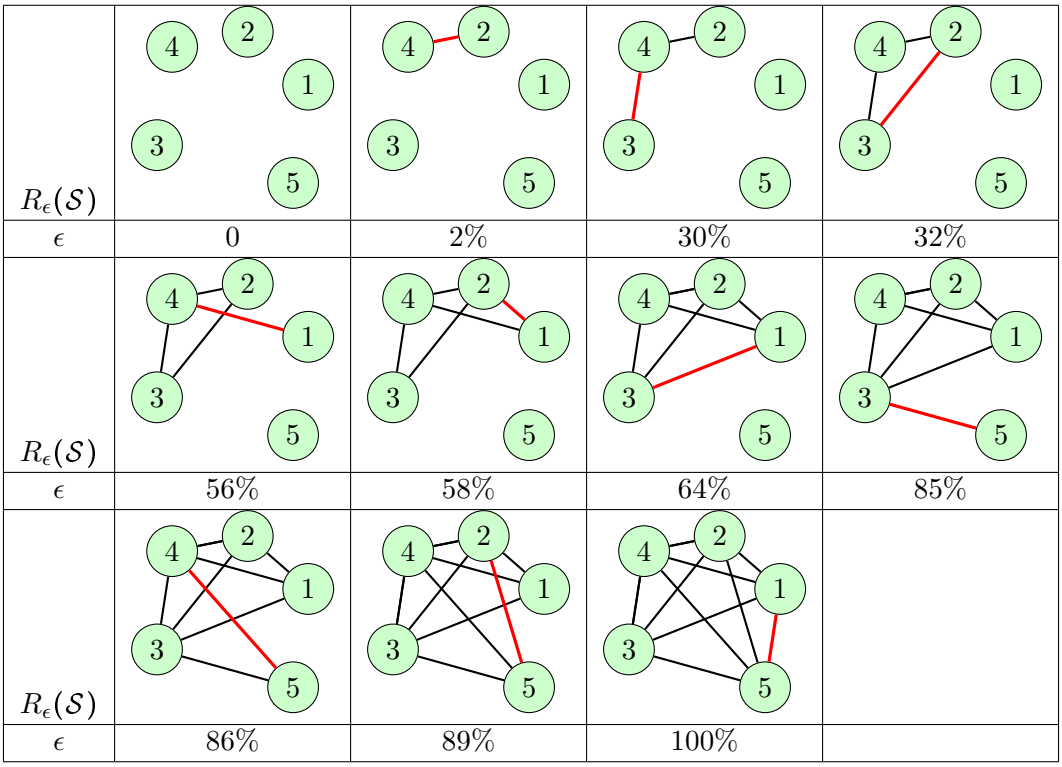


Figure 8.1.5: The filtered simplicial complex associated with score $\mathcal{S}$ from *Laputa: Castle in the Sky*: here each distance from Table 8.1.1 is used as a discretization of time and we represent each complex of the filtration by its associated graph.

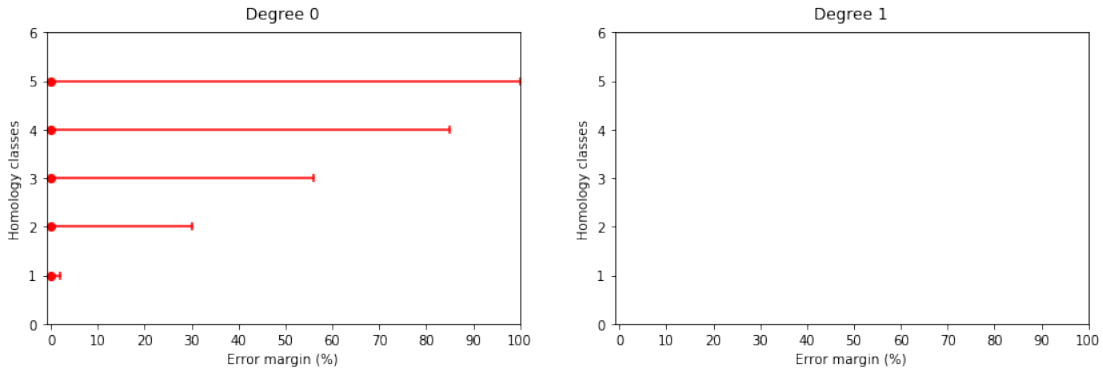


Figure 8.1.6: The family of barcodes associated with score $\mathcal{S}$ from *Laputa: Castle in the Sky* after normalization. Barcode's scale is between 0% and 100%.

Remark 8.1.11. For comparison, we applied the same example with the DFT-approach and get completely different results. The distances between musical bars of score $\mathcal{S}$ from Figure 8.1.4 is presented in Table 8.1.2. The closest bars are $\mathcal{B}_1$ and $\mathcal{B}_5$ (66%) and then each musical bar either join $\mathcal{B}_1$ or $\mathcal{B}_5$. The most distant bars are $\mathcal{B}_2$ and $\mathcal{B}_4$. Therefore, whereas the DFT used as a metric has a strong and consistent musical meaning (Tonnetz, harmonization, classification, etc.), the Hausdorff distance seems to have a completely different musical sense that focuses on the global structure of the score. This new kind of results needs to be exploited, and we give two applications in the following sections: one for a particular musical piece (a lullaby from J. Brahms), and one for a particular musical style (minimalist, with Y. Tiersen).

	$\mathcal{B}_1$	$\mathcal{B}_2$	$\mathcal{B}_3$	$\mathcal{B}_4$	$\mathcal{B}_5$
$\mathcal{B}_1$	0	96%	76%	91%	66%
$\mathcal{B}_2$		0	92%	100%	85%
$\mathcal{B}_3$			0	93%	67%
$\mathcal{B}_4$				0	85%
$\mathcal{B}_5$					0

Table 8.1.2: Table of distances for the musical score $\mathcal{S}$ from *Laputa: Castle in the Sky* using the DFT-distance. The musical bar seems to be gathered according to the number of notes, and not the general structure of the bars.

8.2. Structure of a score and musical texture

8.2.1. General idea

In the light of the previous Example 8.1.10, we want to deeply understand what this new point cloud based on the Hausdorff distance brings to musical analysis: we saw that it deals with the general **structure** of a score, but also to the construction of a musical bar in itself. Here what we call the **musical texture** is deeply linked to the **beats structure** of a musical bar: this means that we give less importance to pitches and focus more on the rhythms. Actually, recalling Definition 8.1.1, a musical bar is a subset of $[0,1]^3$ where two of the three coordinates are related to the **time**. At the end of this first section, we will do some tests to try to normalize the coordinates.

Let us start by considering the well-known lullaby from Johannes Brahms (1833-1897): *Wiegenlied Op.49-4*. Lullabies are usually little music pieces that can be studied in a first

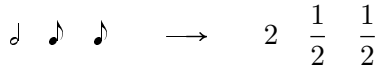
place, in order to understand what musical characteristics we can capture. The score we use for the example is presented in Figure 8.2.1. Here we simplify the original score and especially the accompaniment, which we reduced to the tonic with the corresponding 3-chord, and we focus on the general structure of the melody. Actually, each musical bar contains a part of the melody (first voice) and the second voice is made of the corresponding chord (three notes) with half notes together with the tonic. Notice that the score from Figure 8.2.1 has seventeen musical bars but only fifteen distinct ones ($\mathcal{B}_{10} = \mathcal{B}_{14}$ and $\mathcal{B}_{11} = \mathcal{B}_{15}$).

We computed the filtered complex and the associated family of barcodes according to our new Hausdorff method. This family of barcodes is presented in Figure 8.2.2.

Before starting the analysis, one can notice that the arrangement of the lullaby we made has an accompaniment which is the same in the whole score: actually, each musical bar has a second voice made of one quarter note and the corresponding 3-chords in half notes. It is also the same accompaniment in the original version, but with a different structure. A direct consequence of this fact is that the filtration, and especially the different connected components, (refer to degree 0) will *ignore* the accompaniment, and we will do the analysis with this innuendo. At the end of this section, we will quickly compare this approach with the DFT-distance and see that the accompaniment has more importance in that case (see Section 8.2.3). Also recall that, to have a bijection in the Hausdorff approach, we decided to have three coordinates, and two of them concerned time and rhythms. We will have to make some tests to normalize these coordinates and compare the results (see Section 8.2.2).

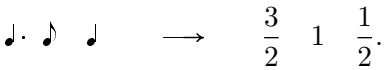
Let us look at the barcode in degree 0 from Figure 8.2.2. There are several levels of analysis depending on the scale we choose to take, and the main idea of persistent homology is to focus on the largest bars (those which *persist*), while the smallest ones can be considered as *noise*. In this case, after 28% of the filtration there are only two connected components that remain: musical bar $\mathcal{B}_1$ is one component and the other is a large dimensional complex where all the other musical bars are connected together. Between 15% and 28%, there are three components: $\mathcal{B}_1$, $\mathcal{B}_{15}$ and the other large complex. We see then that the beginning and the ending of the piece are clearly separated from the rest of the score. Before 11%, there are only small bars in the barcode (noises). We thus need to focus on what happened between 11% and 15% (more precisely at 11%, 12% and 13%): we see that, at a scale of 13%, the filtration looks like in Figure 8.2.3, i.e. the filtered complex made of five connected components.

Now let us analyze this complex. First, since there are exactly five bars in the barcode, the complex is in five components: one sub-complex is made of six musical bars and corresponds to a complete graph, one is a line made of also six musical bars, and there are three isolated vertices. Each component is associated with a **type** of musical bar and these types are successively represented in Figures 8.2.4, 8.2.5 and 8.2.6. The first type forms a complete graph, and most of the vertices correspond to a musical bar that contains one or two half notes (isolated or in a 2-chord) together with two quaver notes. The only one that is slightly different is the musical bar $\mathcal{B}_2$, which contains a dotted quarter note together with a quaver and a quarter note. However, this bar can be linked with this group since, whereas the pitches of the melody are living in the same small space (only seven different pitches for the whole piece: $\{F_4, G_4, A_4, B_{4\flat}, C_5, D_5, F_5\}$), the construction of the beats is almost the same: for $\mathcal{B}_3$, $\mathcal{B}_7$, $\mathcal{B}_9$, $\mathcal{B}_{10}$ and $\mathcal{B}_{11}$, we have



$$\longrightarrow \quad 2 \quad \frac{1}{2} \quad \frac{1}{2}$$

whereas for $\mathcal{B}_2$ we have



$$\longrightarrow \quad \frac{3}{2} \quad 1 \quad \frac{1}{2}.$$

Finally, these six musical bars share almost the same structure (exactly for five of them and almost for one). Since pitches are really closed to each other, the result is the complete graph we get on the right part of Figure 8.1.5.

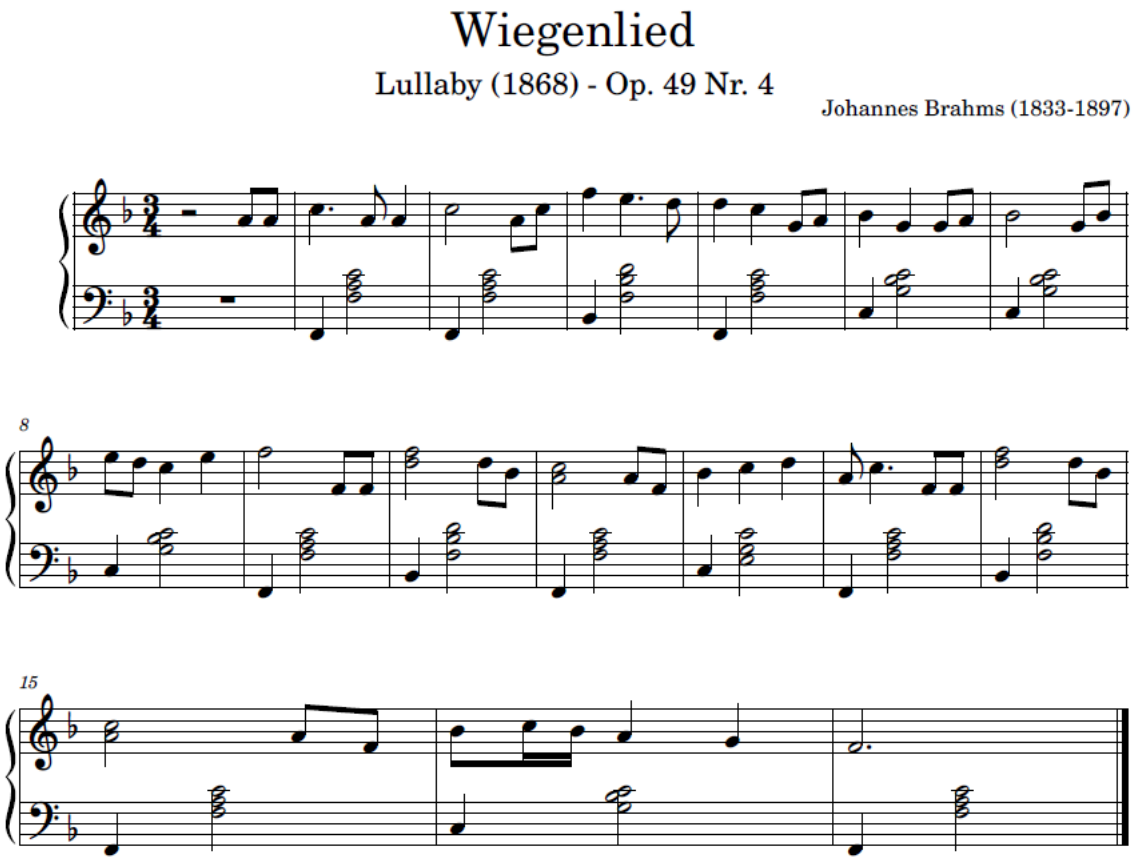


Figure 8.2.1: An arrangement of the original lullaby *Wiegenlied Op.49-4* written by Johannes Brahms in 1868. The original score has been adapted for the analysis: the accompaniment is reduced to the tonic and the corresponding chords, and we are interested in the general structure of the melody.

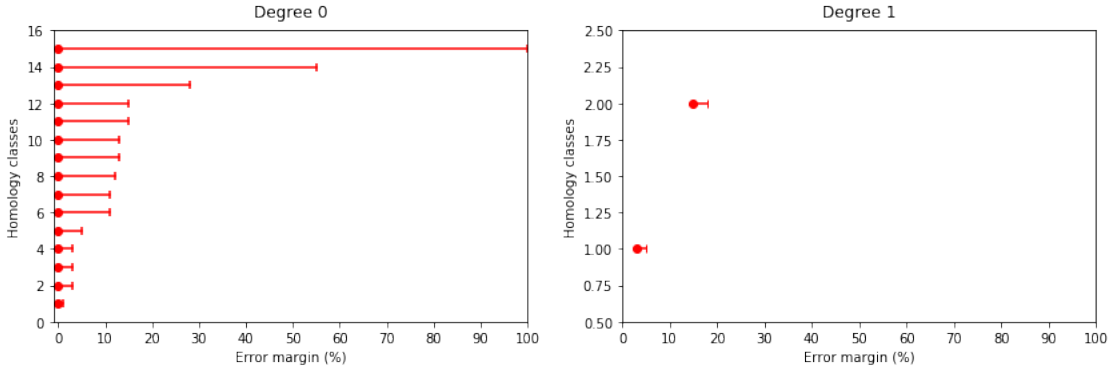


Figure 8.2.2: The associated family of barcodes for *Wiegenlied Op.49-4* in degree 0 (left) and degree 1 (right).

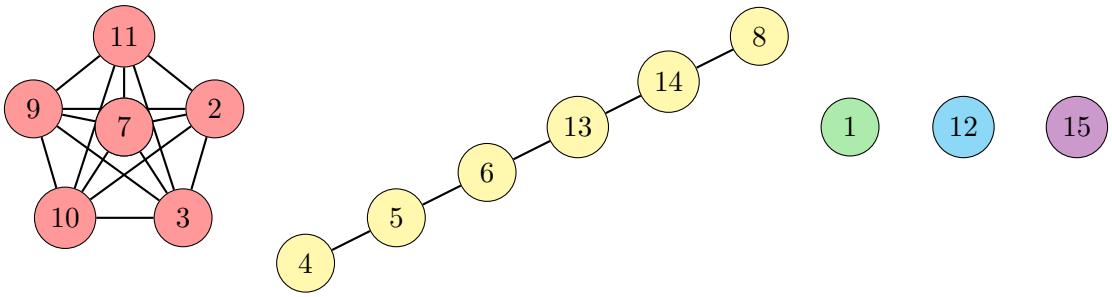
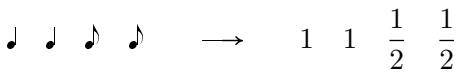
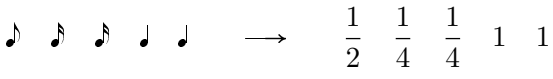


Figure 8.2.3: The filtered complex for *Wiegenlied Op.49-4* at a scale of 13%. There are two main connected components made of the two types of musical bars in the score (a half note with two quaver ones and two quarter notes with two quaver ones), and three made of one vertex, bar $\mathcal{B}_1$ (the beginning), bar $\mathcal{B}_{15}$ (the ending) and bar $\mathcal{B}_{12}$ (a transition made of three quarter notes).

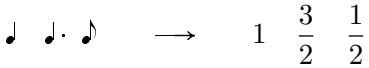
The second type of musical bars is not that obvious, and we see immediately that the graph is simply a straight line, so the bars are only pairwise close. Nevertheless, we understand why these six musical bars can be gathered in only one component: as we did for the first group, we will mostly focus on the rhythms. First, three of these musical contains two quarter and two quaver notes ($\mathcal{B}_5$ and $\mathcal{B}_6$ in this exact order, and $\mathcal{B}_8$ in the reversed one):



This gives us the edge $\{\mathcal{B}_5, \mathcal{B}_6\}$. If $\mathcal{B}_8$ contains two quaver notes and two quarter ones (in that order), $\mathcal{B}_{14}$ has the structure



which seems roughly similar (only the quaver note has been changed for two sixteenths), so we understand the edge $\{\mathcal{B}_8, \mathcal{B}_{14}\}$. Similarly, the beats structure for $\mathcal{B}_{13}$ is close to $\mathcal{B}_5$ and $\mathcal{B}_6$, since it also ends with a quarter note and two quaver ones, and it simply begins with two quaver notes instead of a quarter one. This gives us the edge $\{\mathcal{B}_6, \mathcal{B}_{13}\}$. Finally, $\mathcal{B}_4$ has the same beginning and ending (a quarter and a quaver note) than $\mathcal{B}_5$ and $\mathcal{B}_6$:



so we complete the line with the edge $\{\mathcal{B}_4, \mathcal{B}_5\}$.

Finally, Figure 8.2.6 shows the isolated alone components with the corresponding musical bars: $\mathcal{B}_1$, $\mathcal{B}_{12}$ and $\mathcal{B}_{15}$. Obviously, $\mathcal{B}_1$ is the beginning of the song and $\mathcal{B}_{15}$ the final bar, so we clearly see that these bars do not belong to one of the above group. This is quite satisfying because we can conclude that, with a scaling parameter of 13%, the complex separates the beginning and the ending of the score from the rest of it. Concerning $\mathcal{B}_{12}$, we see that this musical bar is the only one which is made of exactly three quarter notes, so it does not belong to any of the other components. Notice that, at the next step of the filtration (15%), $\mathcal{B}_{12}$ joins the second component by being connected to $\mathcal{B}_5$, $\mathcal{B}_6$ and $\mathcal{B}_8$, the three musical bars containing two quarter notes. At that moment, $\mathcal{B}_2$ is linked with $\mathcal{B}_6$, and that is also interesting because $\mathcal{B}_2$ was the only musical bar of the first group that was slightly different from the five others, while $\mathcal{B}_6$ represents the second group. At that time, $\mathcal{B}_1$ and $\mathcal{B}_{15}$ are in two separated connected component.

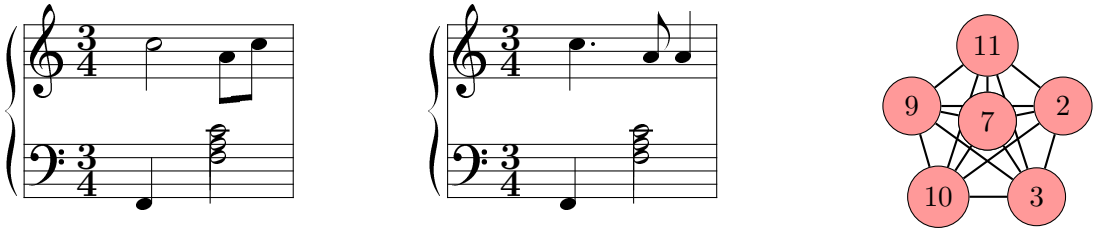


Figure 8.2.4: The first type of musical bar in *Wiegenlied Op.49-4* that form a complete graph (right): most of these bars ($\mathcal{B}_3$, $\mathcal{B}_7$ and $\mathcal{B}_9$) look like the left musical bar, that means a half note with two quaver ones. Musical bars $\mathcal{B}_{10}$ and $\mathcal{B}_{11}$ are made of a chord with two half notes together with two quaver ones, so very similar to the previous case. The musical bar $\mathcal{B}_2$ is made of a dotted quarter note, a quaver and a quarter one.

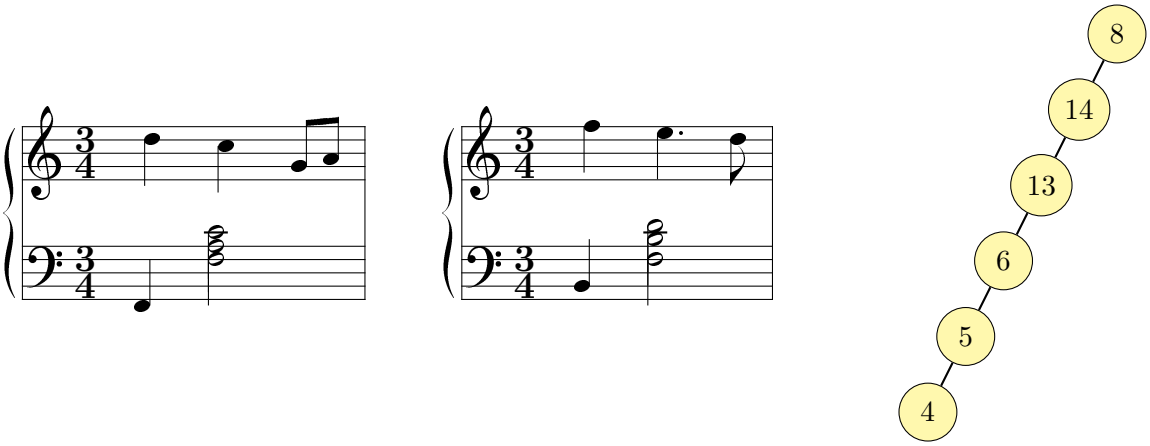


Figure 8.2.5: The second type of musical bar in *Wiegenlied Op.49-4* that forms a straight line: two of these six musical bars share exactly beats structures ($\mathcal{B}_5$ and $\mathcal{B}_6$) and B_8 has the reversed one. $\mathcal{B}_4$, $\mathcal{B}_{13}$ and $\mathcal{B}_{14}$ has same ending or beginning than $\mathcal{B}_5$ or $\mathcal{B}_6$.

Here we give an analysis of what happens in degree 0, and we understand how the musical bars are connected by their general **structure**, mostly using the beats and rhythms. Let us now briefly talk about degree 1: on the barcode from Figure 8.2.2, we see that there are exactly two generators of homology: one that starts at 3% and ends at 5% and the other that starts at 15% and ends at 18%. The first thing to note is that it corresponds to very short bars, especially compared to barcodes from Fourier analysis from other sections (see for instance the barcodes associated with the Tonnetze (see Chapter 5).

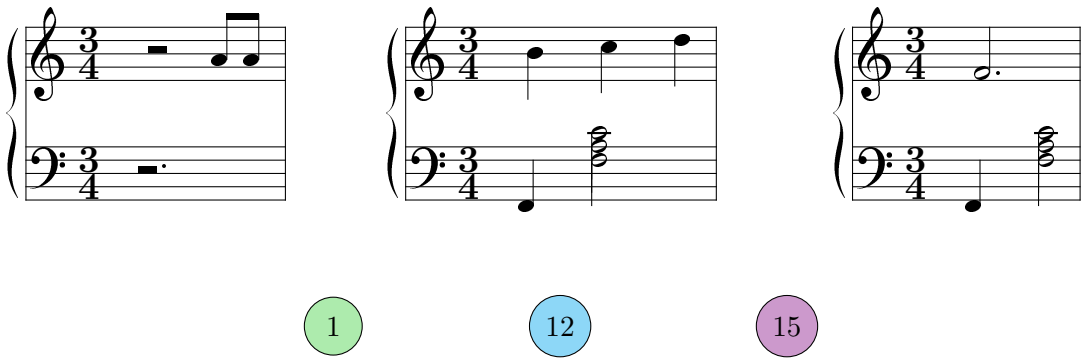


Figure 8.2.6: The three components made of only one vertex in *Wiegenlied Op.49-4* : $\mathcal{B}_1$ and $\mathcal{B}_{15}$ are respectively the beginning and the end of the score, which are separated from the rest of the score, and $\mathcal{B}_{12}$ is the only one which contains exactly three quarter notes.

We represented these two one-dimensional cycles in Figure 8.2.7. The first one (left) is a cycle that linked musical bars from the first group that share the *exact* same type: indeed, the musical bars $\mathcal{B}_3$ and $\mathcal{B}_7$ are made of one half note and two quaver ones, while $\mathcal{B}_{10}$ and $\mathcal{B}_{11}$ are made of a 2-chords with half notes and two quaver ones.

One the other hand, the second cycle (right) has length 5 and is made of musical bars from the second type together with $\mathcal{B}_{12}$, which was one of the isolated vertices. Unlike the first cycle, this one links $\mathcal{B}_5$ with its reverse one $\mathcal{B}_8$ and $\mathcal{B}_{13}$ with $\mathcal{B}_{164}$, which are musical bars having specific structure in this group.

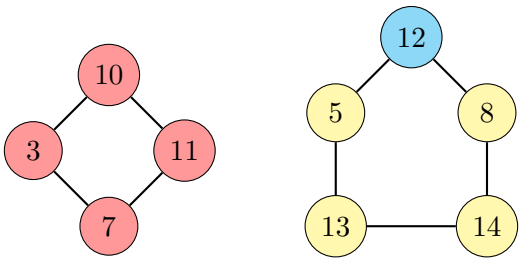


Figure 8.2.7: One-dimensional cycles from *Wiegenlied Op.49-4* that appear at a scale of 3% (left) and 15% (right).

To conclude with this first example, we saw that the Hausdorff approach seems to have a particular musical meaning concerning the structure of the musical bars and the score in itself. With *Wiegenlied Op.49-4*, we saw that the degree 0 seems to gather the different musical bars and classify them according to their beats structure. We will change the coordinate weights of each note from our Definition 8.1.1.

8.2.2. Change coordinate weights

The Hausdorff point clouds we define in 8.1.9 are built in $\mathbb{R}^3$, and each note has two coordinates concerning time and only one concerning pitches. We might want to see what happens if we put a **weight** on these coordinates, that means a triplet $(a, b, c) \in \mathbb{R}^3$ that we apply on a note $(n_1, n_2, n_3) \in [0, 1]^3$ with the termwise product (an_1, bn_2, cn_3).

❋ **Weight** $(\frac{1}{2}, \frac{1}{2}, 1)$**:** With this first triplet, we want to give equal importance to time and frequencies. The result is the barcode from Figure 8.2.8: we see that it is roughly similar to the original barcodes. Indeed, we have short bars (noises) until 15%, only two components at 32% that corresponds to bars $\mathcal{B}_1$ and $\mathcal{B}_{15}$ so this aspect is preserved. Between 15% and 32%, we have several levels of filtration and more precisely, at 16%, the complex looks like in Figure 8.2.9. We see on this graph that the first group is preserved but bar $\mathcal{B}_2$ is only linked to $\mathcal{B}_{11}$ instead of every vertices. This is probably due to the fact that $\mathcal{B}_{11}$ has the most common pitches with $\mathcal{B}_2$, and here we give more weight to pitches. On the other hand, the second group is split in two parts. Musical bars $\mathcal{B}_1$, $\mathcal{B}_{12}$ and $\mathcal{B}_{15}$ are still alone in their respective components. Concerning degree 1, we also have only two components, that are exactly the same as in the original analysis, and these one-dimension cycles appear almost at the same time (from 7% to 10% for the first one and from 23% to 25% for the second one). In conclusion, the analysis is almost the same, with some small details that change inside each component but the general structure is preserved.

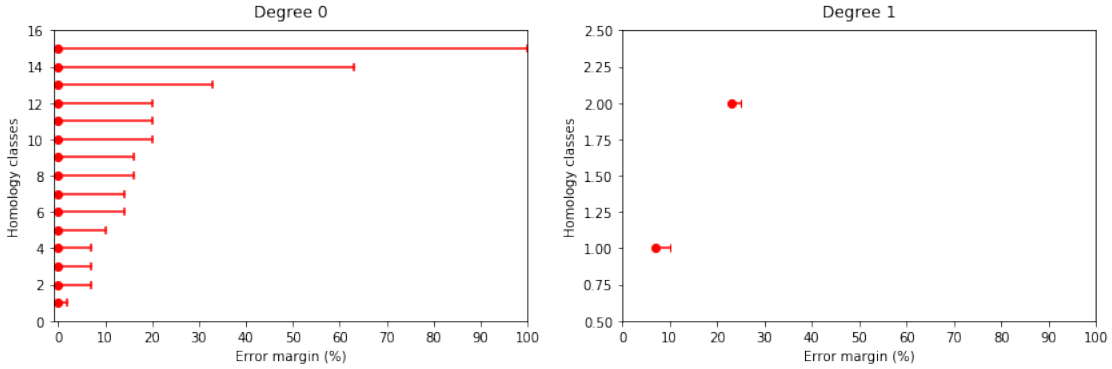


Figure 8.2.8: The family of barcodes for *Wiegenlied Op.49-4* obtained with the Hausdorff distance by adding a weight $(\frac{1}{2}, \frac{1}{2}, 1)$.

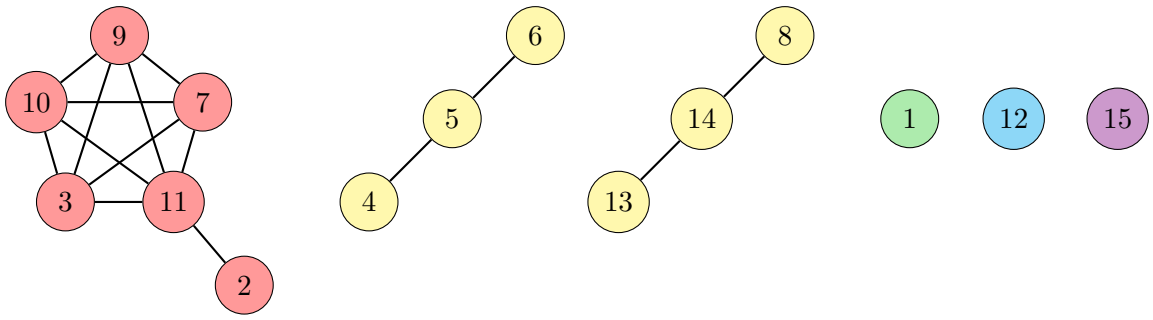


Figure 8.2.9: The filtered complex for *Wiegenlied Op.49-4* at a scale of 16%. The point cloud here is computed by adding a weight $(\frac{1}{2}, \frac{1}{2}, 1)$.

❋ **Weight** $(1, 0, 1)$**:** In this paragraph, we try a different approach by building the same point cloud as we did for the Fourier analysis, i.e. we simply keep the onsets and the pitches. The resulting barcodes are given in Figure 8.2.10. We can immediately notice that the filtration stays longer in two components (until 91% here against 56% for the basic analysis and 63% for the weight $(\frac{1}{2}, \frac{1}{2}, 1)$). The musical bar that stays outside the large complex is still the beginning $\mathcal{B}_1$, the only one which does not have an accompaniment. The musical bars $\mathcal{B}_{12}$ and $\mathcal{B}_{15}$ are still two isolated components during a while (respectively until 17% and 19%). For comparison with the previous weight, we chose to represent the filtration with a scaling parameter of 12%, so the complex is in six components, as shown in Figure 8.2.11.

In this Figure, if $\mathcal{B}_8$, $\mathcal{B}_{13}$ and $\mathcal{B}_{14}$ still represent a component, we see that $\mathcal{B}_4$, $\mathcal{B}_5$ and $\mathcal{B}_6$ have joined the first group: for instance, $\mathcal{B}_6$ is now linked with $\mathcal{B}_7$, and we can observe that these two musical bars share the same accompaniment, so this is now taken into account. Concerning degree 1, there is only one remaining bar which starts at 6% and stops at 8%, and it is given by the same first one as for the previous analysis:

$$\{\mathcal{B}_3, \mathcal{B}_7\} + \{\mathcal{B}_7, \mathcal{B}_{11}\} + \{\mathcal{B}_{11}, \mathcal{B}_{10}\} + \{\mathcal{B}_{10}, \mathcal{B}_3\}.$$

In conclusion, the biggest change here is the accompaniment that is now taken into consideration, and the fact that we not only focus on the melody anymore. These results are roughly closed to the DFT-approach (see the next paragraph), so for the rest of the analysis and especially the paragraph on the minimalist music 8.3, we will still keep the classical distance, without additional weights. In fact, what is interesting now is the way the Hausdorff distance captures the global structure of the score. This works particularly well when applied to minimalist music, since the accompaniment has the same rhythmic structure regardless of pitches. We conclude this section by comparing the results with those obtained using the DFT-approach on the same score.

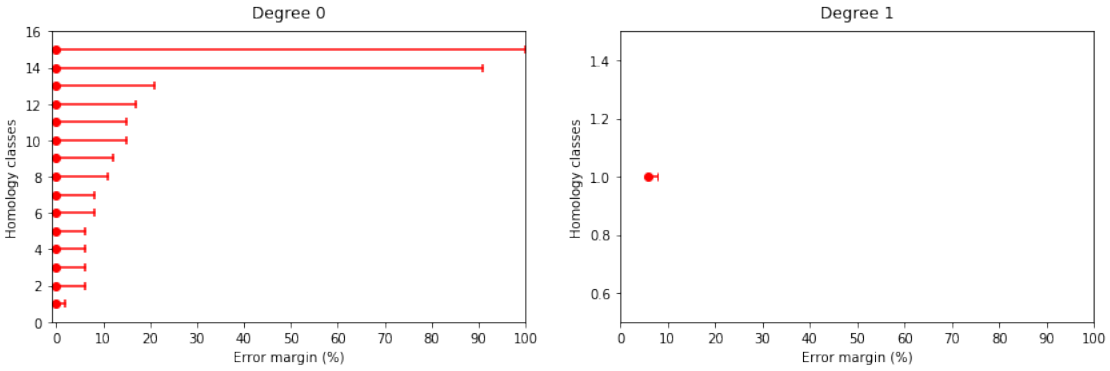


Figure 8.2.10: The family of barcodes for *Wiegenlied Op.49-4* obtained with the Hausdorff distance by adding a weight $(1, 0, 1)$.

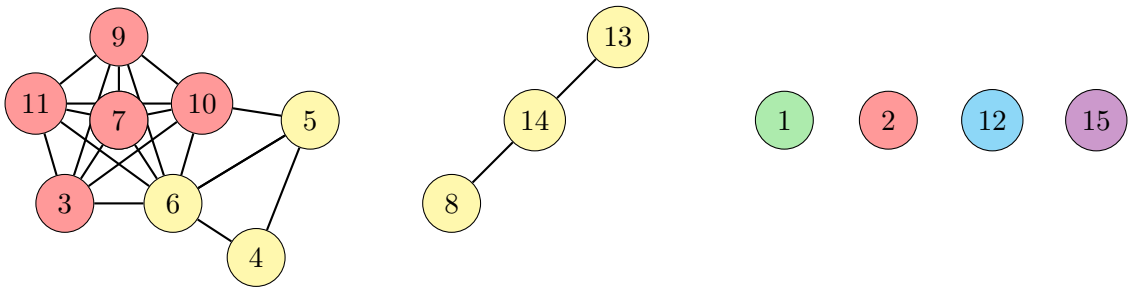


Figure 8.2.11: The filtered complex for *Wiegenlied Op.49-4* at a scale of 12%. The point cloud here is computed with an adding weight of $(1, 0, 1)$.

8.2.3. Comparison with the DFT-distance

To conclude with the lullaby from Johannes Brahms, we simply give a comparison with the DFT-approach: the resulting barcodes are given in Figure 8.2.12. For degree 1, there is also one generator, as it was the case for the weight $(1, 0, 1)$. However, this new one-dimensional cycle is now given by

$$\{\mathcal{B}_1, \mathcal{B}_4\} + \{\mathcal{B}_4, \mathcal{B}_{10}\} + \{\mathcal{B}_{10}, \mathcal{B}_7\} + \{\mathcal{B}_7, \mathcal{B}_1\}$$

which is different from the previous cycles. In the same way, the barcode in degree 0 looks not the same from the one obtained with the Hausdorff distance: in fact, it seems more homogeneous, all the musical bars seem to come together from 37% and with regular scale. The complex stays in two components until 79%, and these components are given by the edge $\{\mathcal{B}_4, \mathcal{B}_{10}\}$ on the one hand (two bars that share the same accompaniment, and not the beginning and the ending of the score), and the rest of the complex on the other.

In order to compare with the Hausdorff approach, we now try to find a scale where the complex is in several components and see how these groups are connecting. We represent the complex at 58% of the filtration, as shown in Figure 8.2.3 . The complex is in six components, and there are some obvious differences with the previous analyses, but also some similarities: here $\mathcal{B}_1$ and $\mathcal{B}_{12}$ are isolated vertices, but $\mathcal{B}_{15}$ is included into a graph, while $\mathcal{B}_8$ is now a component itself. Secondly, the two first components of the Hausdorff analysis are now mixed together. If we focus on each component, we might see that they are now gathered according to their accompaniment, and not only the global structure anymore. For instance, edge $\{\mathcal{B}_4, \mathcal{B}_{10}\}$ represents the only two musical bars which accompaniment is a $B\flat$ together with 3-chords $\{F, B\flat, D\}$. The triangle $\{\mathcal{B}_6, \mathcal{B}_7, \mathcal{B}_{14}\}$ represents the three musical bars which accompaniment is a C together with 3-chords $\{G, B\flat, C\}$.

Therefore, we understand that the DFT-approach seems to focus more on the accompaniment, and even more than the Hausdorff did with weight $(1, 0, 1)$. Therefore, we decided to keep working with the classical Hausdorff distance: indeed, it seems to be the metric that captures the global structure of a score as well as possible in our study. However, if we rather want to take the accompaniment into account, we can still consider the DFT-distance, which seems more appropriate in this context, according to what has just been done.

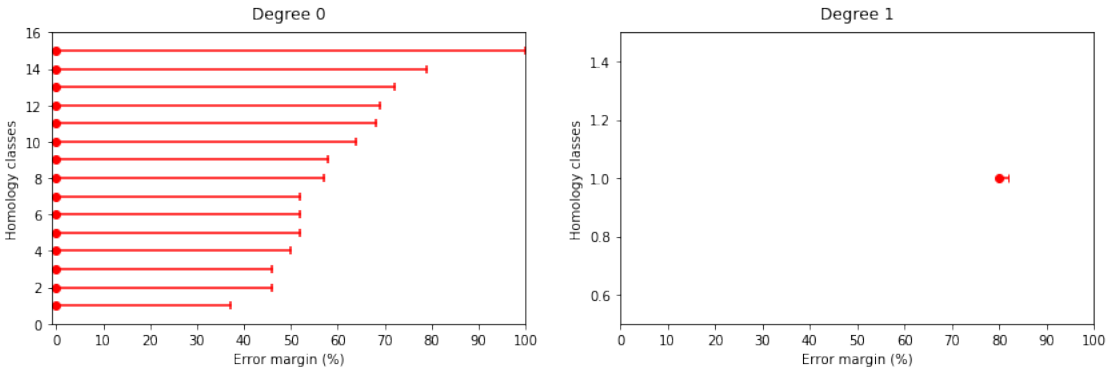


Figure 8.2.12: The barcodes family for *Wiegenlied Op.49-4* obtained with the DFT-method in degree 0 (left) and degree 1 (right).

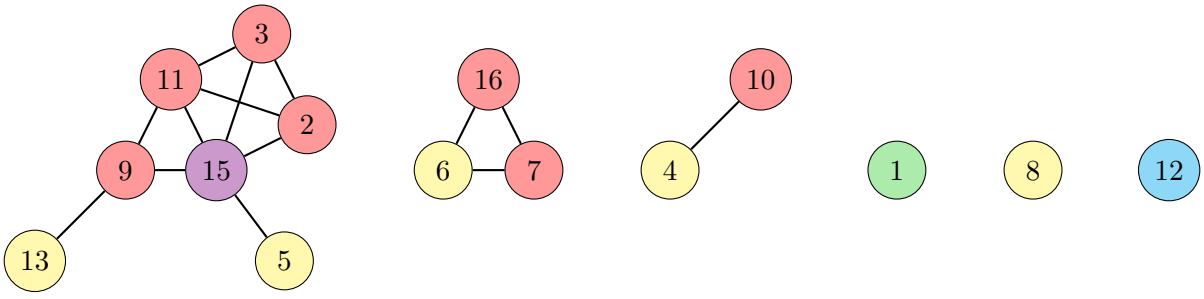


Figure 8.2.13: The filtered complex for *Wiegenlied Op.49-4* at a scale of 58%. The point cloud here is computed with the DFT-method.

8.3. Application to minimalist music: Yann Tiersen

Yann Tiersen (born June 23, 1970) is a French musician and composer known among other things for the soundtrack of the film *Le Fabuleux Destin d'Amélie Poulain*, directed by Jean-Pierre Jeunet (2001). The musical style of Yann Tiersen is often classified as **minimalism** and **repetitive music**: in fact, many of his songs are composed over the same accompaniment by making variation of one or several main themes. These are the main reasons why we chose to study this composer in the context of the Hausdorff analysis, in light of the results obtained for the lullaby of Johannes Brahms in the previous sections. Our aim is thus to see how this can be applied on different music pieces from this particular minimalist musical style.

Here we are going to study three important pieces of this style: *Comptine d'un autre été: l'Après-midi* (2001), *Comptine d'été No. 2* (1996) and *La Dispute* (1998).

8.3.1. Comptine d'un autre été: l'Après-midi

The score has fifty-three musical bars, but seems some of them are repeated, it only contains thirty-nine distinct bars. It is also made of two main parts: actually, the music has three different themes in the first part that are played again one octave higher in the second one. These three themes are presented in Figure 8.3.1. Notice that there are some repetitions of these themes in the original score, but here we suppress them in order to work with distinct musical bars only. Moreover, all the melody is constructed over four musical bars that are repeated and which constitute the musical accompaniment from Figure 8.3.2. Finally, the ending bar $\mathcal{B}_{39}$ contains only one chord Em played with whole notes, as shown in Figure 8.3.3.

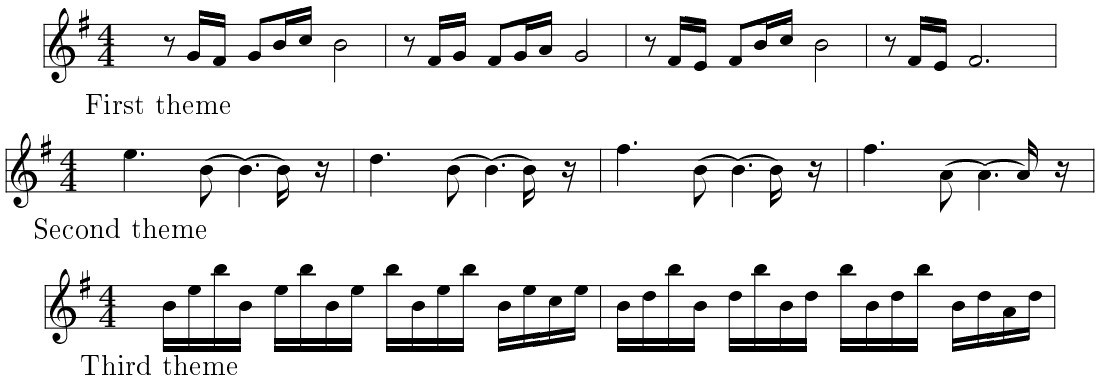


Figure 8.3.1: Part of each theme of *Comptine d'un autre été: l'Après-midi*. The first one goes from $\mathcal{B}_5$ to $\mathcal{B}_8$ and is repeated one octave higher from $\mathcal{B}_{22}$ to $\mathcal{B}_{25}$. The second goes from $\mathcal{B}_9$ to $\mathcal{B}_{12}$, then is repeated with extra notes from $\mathcal{B}_{13}$ to $\mathcal{B}_{16}$ and the whole eight bars are played one octave higher from $\mathcal{B}_{26}$ to $\mathcal{B}_{33}$. The third one goes from $\mathcal{B}_{17}$ to $\mathcal{B}_{21}$ and is repeated one octave higher from $\mathcal{B}_{34}$ to $\mathcal{B}_{38}$, with a slight change in $\mathcal{B}_{38}$ to bring us to the end of the piece.

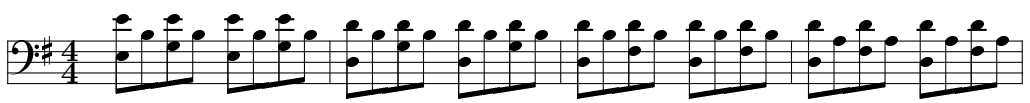


Figure 8.3.2: The musical accompaniment of *Comptine d'un autre été: l'Après-midi*. These four bars consists of four arpeggiated chords $Em - G - B - D$ and are played once at the beginning of the score without any melody, from $\mathcal{B}_1$ to $\mathcal{B}_4$.

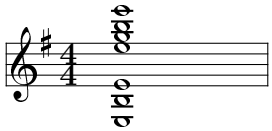


Figure 8.3.3: The end of the piece of *Comptine d'un autre été: l'Après-midi* which corresponds to the last bar of the score (the $39th$). We recover the main chord Em played with whole note.

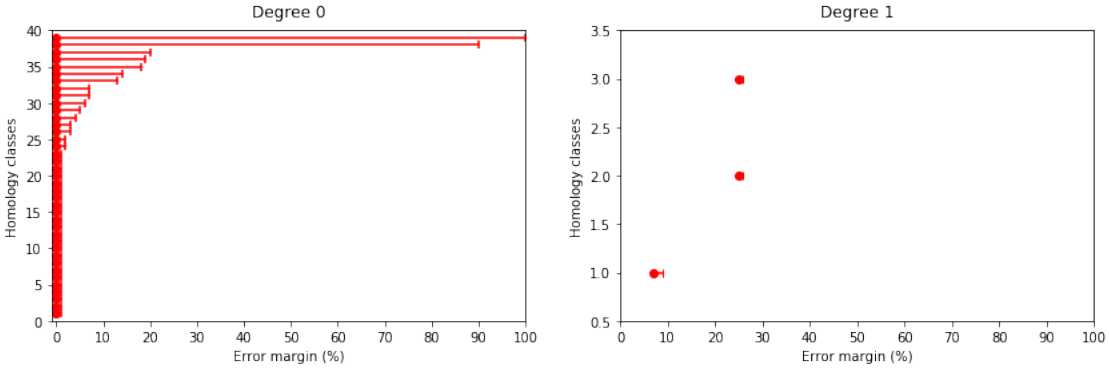


Figure 8.3.4: The associated family of barcodes with *Comptine d'un autre été : L'Après-midi* in degree 0 (left) and degree 1 (right).

Let us look at the barcode in degree 0 from Figure 8.3.4: in our case, there are two bars that stand out when we take a scale larger than 21%, that means that the corresponding complex has only two connected components. One of them corresponds to the last musical bar $\mathcal{B}_{39}$ of the score, which only consists of the final chord Em played with whole notes, and the other is a large dimensional complex where all the musical bars are connected together. This first analysis shows that the barcode in degree 0 separates the end from the rest of the piece, which is a start. For a scaling parameter smaller than 8%, there are only small bars so we ignore them and consider as noise. Between 8% and 21%, there are five, six seven classes that seem to last and more precisely, we found that at a scale of 14%, the associated complex in the filtration looks like in Figure 8.3.5, which is really remarkable: actually, there are six connected components and if we look at the vertices, we see that each one corresponds to a theme of the song, except for $\mathcal{B}_8$ and $\mathcal{B}_{28}$ which have a slight different structure than the rest of the first theme (see the fourth musical bar in the first score of Figure 8.3.1). We may then conclude that, with a well-chosen scale, the barcode in degree 0 captures the structure of this piece by separating its different themes in the associated complex.

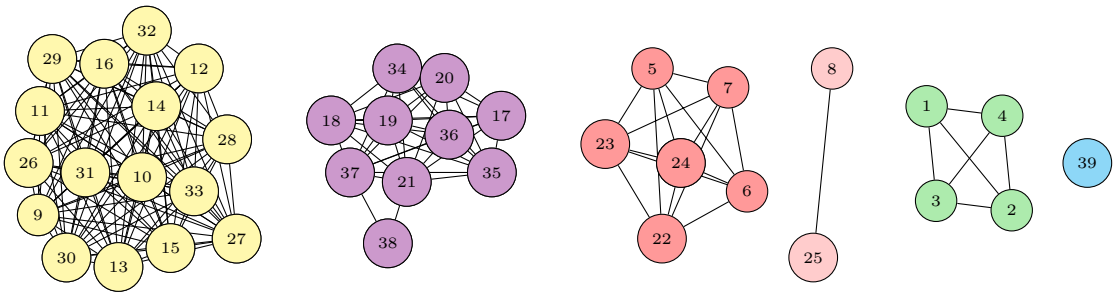


Figure 8.3.5: The associated complex of *Comptine d'un autre été: l'Après-midi* at a scale of 14%: each component characterizes a theme of the piece.

On the other hand, the barcode in degree 1 displays three different one-dimensional cycles, which are presented in Figure 8.3.6. Note that some edges of these cycles connect musical bars of one given theme to the same one octave higher, but not systematically and for now we are not able to interpret these cycles musically. Moreover, these cycles have very short lifetimes in the barcode (only 1% and 2%), so they can be considered as noise. For these reasons and as we already did for the Lullaby, we will mainly focus on degree 0 in the following analyses.

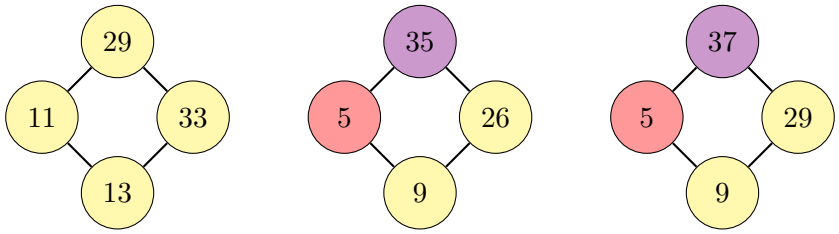


Figure 8.3.6: One-dimensional cycles from *Comptine d'un autre été : L'Après-midi* that appear at a scale of 8% (left) and 26% (middle and right).

8.3.2. Comptine d'été No.2

This song is roughly similar to the previous "comptine": the score has sixty-five musical bars but only thirty distinct ones and is built on an accompaniment made of four distinct musical bars that are presented in Figure 8.3.8. This accompaniment is played without melody on the eight first bars of the score: then, there are several themes that follow one after the other and which are briefly presented in Figure 8.3.7. By counting with distinct bars, the first goes from $\mathcal{B}_5$ to $\mathcal{B}_{12}$, the second from $\mathcal{B}_{13}$ to $\mathcal{B}_{18}$, the third from $\mathcal{B}_{19}$ to $\mathcal{B}_{25}$ and the last one from $\mathcal{B}_{26}$ to $\mathcal{B}_{29}$. As for the *Comptine d'un autre été: l'Après-midi*, the song ends with a chord of conclusion in musical bar $\mathcal{B}_{30}$, as presented in Figure 8.3.9.

According to the first analysis we did with *Comptine d'un autre été: l'Après-midi*, we are expected here to find a scale for which the filtered complex would be in six components: one for each theme, one for the four musical bars of accompaniment and one for the conclusion chord. We computed persistent homology and the family of barcodes is presented in Figure 8.3.10.

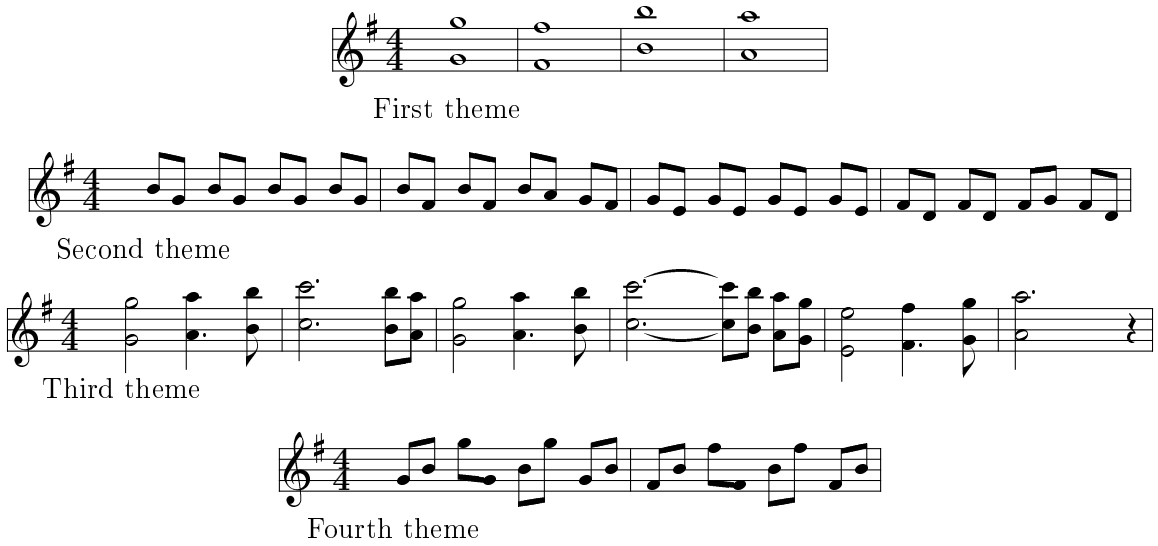


Figure 8.3.7: Part of each theme of *Comptine d'été No. 2*. The first one goes from $\mathcal{B}_5$ to $\mathcal{B}_{12}$, while the second goes from $\mathcal{B}_{13}$ to $\mathcal{B}_{18}$ and which is close to the last theme, at least in terms of beats structure. The third theme goes from $\mathcal{B}_{19}$ to $\mathcal{B}_{25}$.

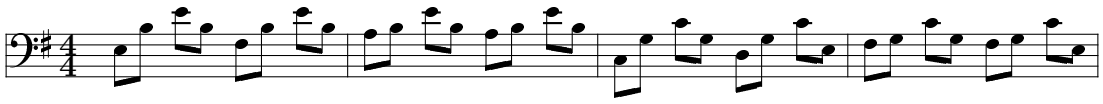


Figure 8.3.8: The musical accompaniment of *Comptine d'été No. 2*. These four bars consist of four arpeggiated chords and are played once at the beginning of the score without any melody, from $\mathcal{B}_1$ to $\mathcal{B}_4$.

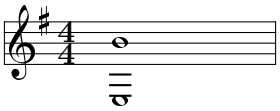


Figure 8.3.9: The end of the piece of *Comptine d'été No. 2* which corresponds to the last bar of the score $\mathcal{B}_{30}$, made of the 2-chords $\{E_3, B_4\}$ played with whole notes.

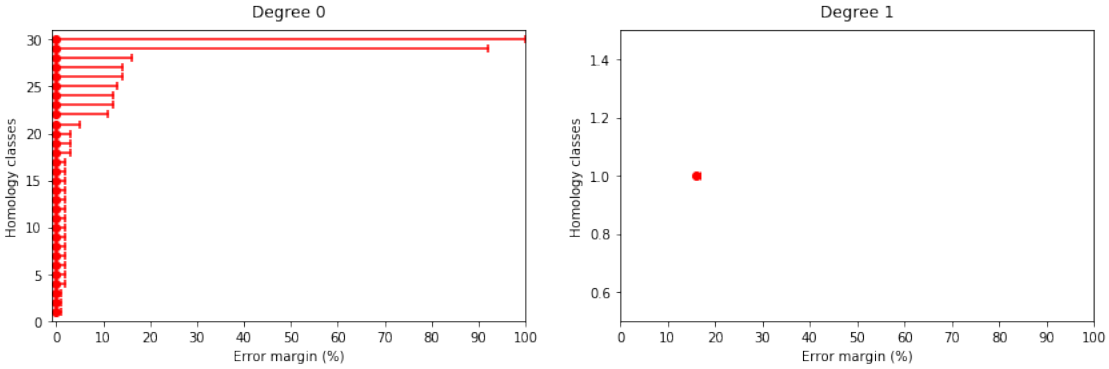


Figure 8.3.10: The associated family of barcodes for *Comptine d'été No. 2* in degree 0 (left) and degree 1 (right).

We can immediately notice that this family of barcodes is roughly similar to the one from *Comptine d'un autre été: l'Après-midi*. Actually, in degree 1, we see only one generator which is a very short bar that corresponds to a one-dimensional cycle. As we already saw in the previous example and before, this kind of cycle is rather difficult to interpret musically and also because it represents a really short bar in the barcode, we will consider it as noise and thus ignore it.

Let us focus on the barcode in degree 0. As for *Comptine d'un autre été: l'Après-midi*, we only have two bars left when we take a scale larger than 16% here: one of them corresponds to the last musical bar $\mathcal{B}_{31}$ and the other is the large dimensional complex with all the other interconnected musical bars of the score. Here we find that, at a scale of 5%, which is rather a small parameter, the complex looks like in Figure 8.3.11. On this complex, we recover several connected components (exactly eight by counting the isolated vertices), and we see that each group represents again a theme in the score. In fact, the largest complex (on the left) corresponds to the second theme together with the last one, the two being exactly the same in terms of beats structure (same accompaniment and eight quaver notes for each musical bar of the melody). The second complex on the left is the first theme, while the completed square is the musical bars of accompaniment and the isolated right vertex is the ending bar.

Concerning the middle complex, which has several components itself, the six vertices belong to the third theme: in fact, if the musical bars that compose it are closed, they are still slightly different at least compared to the rest of the score. That is why they would not be linked before some of the other components are glued together: when these six components are connected, the first and second theme will also be gathered, so we prefer to display the complex with this scale.

However, all these six vertices belong to the same theme anyway, so we still recognize this part of the song. In conclusion, we find back in this degree 0 analysis, with a well chosen scale, a moment when the filtered complex classifies the different part of the song.

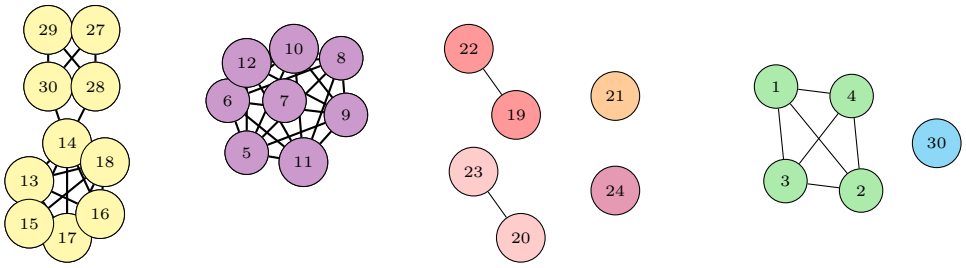


Figure 8.3.11: The associated complex of *Comptine d'été No. 2* at a scale of 5%: each component characterizes a theme of the piece.

8.3.3. La Dispute

This last score has ninety-six musical bars but only fifty-four distinct ones. This music piece is quite different from the two "comptines" we studied previously: in fact, here the song does not start with the same four musical bars of accompaniment but it begins directly with a waltz in two voices. The other main difference is that the accompaniment is not the same during the whole piece: more precisely, the song is divided in three main parts, a waltz (from $\mathcal{B}_1$ to $\mathcal{B}_{48}$, see Figure 8.3.12), a bridge with arpeggio accompaniment (from $\mathcal{B}_{49}$ to $\mathcal{B}_{72}$, see Figure 8.3.13) and an octave passage (from $\mathcal{B}_{73}$ to $\mathcal{B}_{94}$, see Figure 8.3.14). The song ends with two musical bars: $\mathcal{B}_{95}$ brings to the final Cm chord with half dotted notes, which is shown in Figure 8.3.15. According to this description, we would expect here to find a scale for which the filtered complex is in four or five components. The family of barcodes is presented in Figure 8.3.16. As we did for the two previous songs, we only focus here on barcode in degree 0.

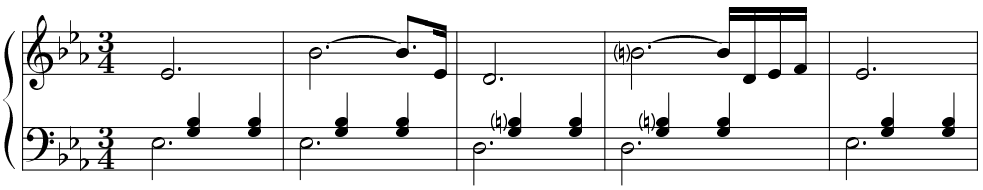


Figure 8.3.12: The beginning of the first theme of *La Dispute*: it goes from musical bars $\mathcal{B}_1$ to $\mathcal{B}_{13}$ and is then repeated by adding the same melody voice one octave higher, from musical bars $\mathcal{B}_{14}$ to $\mathcal{B}_{24}$.

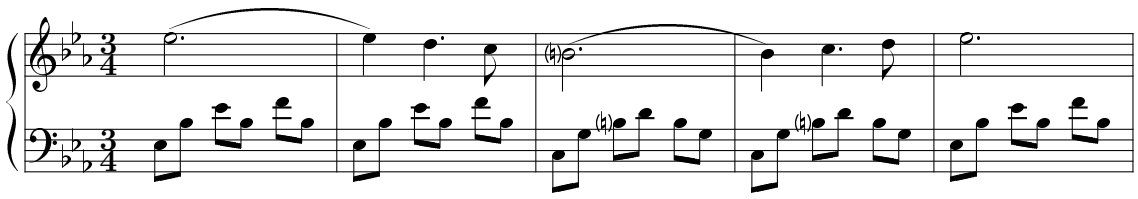


Figure 8.3.13: The beginning of the second theme of *La Dispute*: it goes from musical bar $\mathcal{B}_{25}$ to $\mathcal{B}_{39}$.

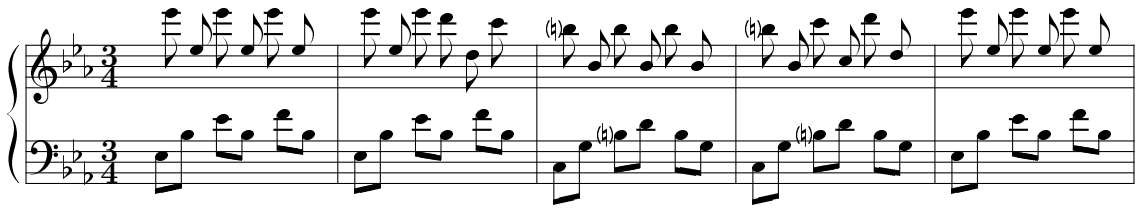


Figure 8.3.14: The beginning of the third theme of *La Dispute*: it goes from musical bar $\mathcal{B}_{40}$ to $\mathcal{B}_{53}$.

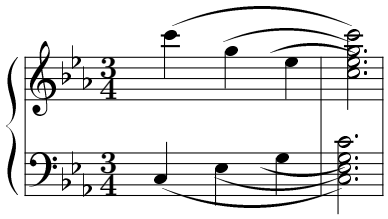


Figure 8.3.15: The end of the piece of *La Dispute* which corresponds to the last two bars of the score $\mathcal{B}_{54}$ and $\mathcal{B}_{55}$, made of a bridge and a Cm chord played with half dotted notes.

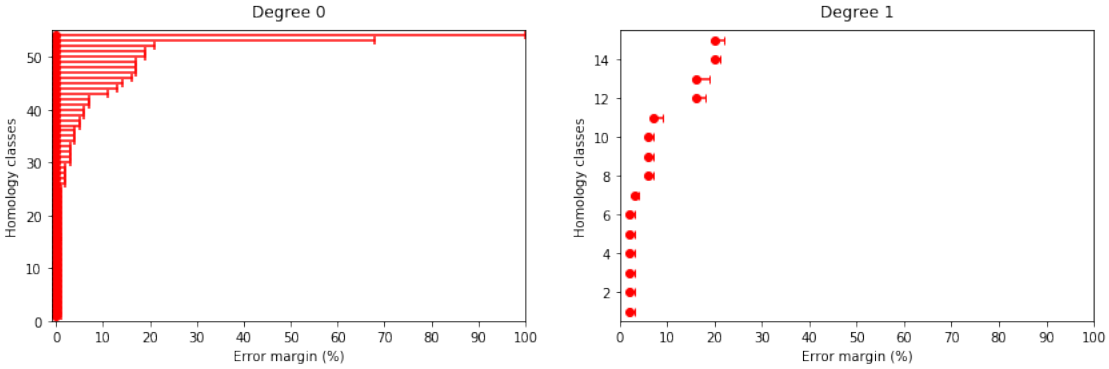


Figure 8.3.16: The associated family of barcodes for *La Dispute* in degree 0 (left) and degree 1 (right).

These two barcodes are again roughly similar than the ones obtained for the two other songs from Yann Tiersen: we recover the two main components after a certain level of filtration, here with a scaling parameter of 21%, and noises for scales smaller than 11%. Between these two moments, and more precisely at a scale of 16%, we see that the filtration looks like in Figure 8.3.17, where the complex is divided in eight components instead of the expected five complexes. However, this moment of the filtration is quite relevant of the general structure of the score: in fact, the complex on the very left of the figure corresponds to the third part of the score. The second complex just next to it, which is actually in two components, is the middle theme of the piece: indeed, most of the musical bars contains a half dotted note or at least a half one with some quaver notes, except for two musical bars ($\mathcal{B}_{26}$ and $\mathcal{B}_{28}$) whose beats structure is given by

$$\text{♩ ♩. ♪} \quad \longrightarrow \quad 1 \ \frac{3}{2} \ \frac{1}{2}$$

Because of this slight difference (they still share the same arpeggio accompaniment), these two musical bars are isolated from the complex. On the right, we see two alone vertices, that are the two final musical bars $\mathcal{B}_{54}$ and $\mathcal{B}_{55}$, as expected. Then, we have this middle complex

which represents the first waltz part of the song and which is actually in three components: two exact complexes and two edges. In this waltz part, there are two types of musical bars that we differentiate with the melody: the one with a half dotted note and the other with a half note together with a eighth or sixteenth notes. Here we recover these two types of musical bars in each one of the biggest middle complexes, which is satisfying. Finally, the isolated edge is given by musical bars $\mathcal{B}_{12}$ together with the same but one octave higher $\mathcal{B}_{23}$, whose beats structure is different from the rest since it is given by three quarter notes. In conclusion, with a well chosen scale, the degree 0 provides a moment when the filtered complex classifies the different part of the song.

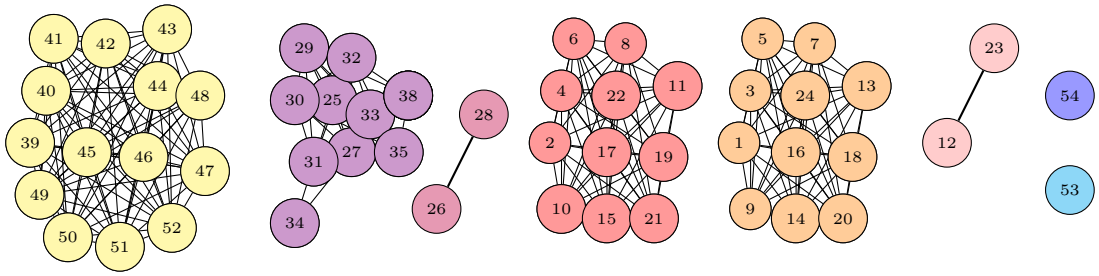


Figure 8.3.17: The associated complex of *La Dispute* at a scale of 16%: each component characterizes a theme of the piece.

Conclusion and perspectives for future research

In this thesis we have tried to answer general questions that are fundamental to topological analysis applied to musical data: how should we associate a filtered complex with a given piece of music? What can such a representation contribute to musical analysis? How much does it help us to understand music? For example, is it able to classify several different styles? In this particular context, we focused on the symbolic representation of music, as opposed to audio and signal, and such a representation is given by MIDI files.

In the introduction, we saw that the question of how to associate a filtration with a musical score is central to the symbolic part of the MIR community, and there are many choices to be made. In our work, we decided to define a score as an non-ordered finite set of distinct musical bars: therefore, we simply zoom in on each part of the piece and keep the information given by pitches and onsets. More precisely, a musical bar is a subset of $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, where t and p are units of time and pitch respectively, and a note played in the piece is characterized by its onset and its pitch.

This representation allows us to use the DFT in order to associate a matrix of Fourier coefficients with each musical bar of a given score, and thus to compare two musical bars using the so-called DFT-distance. In this way, it is possible to extract a point cloud from a musical score, and filtration comes naturally using the Vietoris-Rips method. We created a new model to associate a topological fingerprint to any musical score, given by the corresponding family of barcodes.

As a direct application of this model, our first aim was to generalize some classical results about the DFT in a musical context, such as Lewin's Lemma or Hexachordal Theorems (see Section 2). In particular, since we defined a metric space given by the set of all musical bars in $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$ together with the DFT-distance $(\mathfrak{B}_{t,p}, \mathrm{d}_{\mathrm{DFT}})$, a natural question was to look at the isometry group for this metric. By generalizing work on the dihedral group applied to basic musical structures, we found that rotations (musical transpositions), reflections (musical inversions) and homotheties are all isometries for our DFT-distance. A natural question that we plan to consider in our future work is to look at other isometries for this metric, and obviously this is a tricky question considering the number of elements in $\mathfrak{B}_{t,p}$.

In this conclusion and opening chapter, we want to give an overview of the work done in this thesis: more precisely, for each musical application, we mention the main results and propose some research axes for future work.

Inverse problem: applications on scales and chords

In the first two Parts I and II of this manuscript, we introduced a new method for extracting a filtered complex from a score. To support this approach, we wanted to consider the inverse problem, that is, to apply the DFT together with persistent homology to artificially constructed musical scores and show that the DFT is a reasonable choice of metric.

★ The results

The musical scores we built were first given by single notes in a musical bar: for instance, the chromatic scale is a score with twelve distinct musical bars, each containing only one element in $\mathbb{Z}/1\mathbb{Z} \times \mathbb{Z}/12\mathbb{Z}$. When the scale is well chosen, as in the case of diatonic scales, the DFT seems to recognize some musical features, such as the circle of thirds, which appears as a one-dimensional cycle in H_1, as shown in Figure 1.

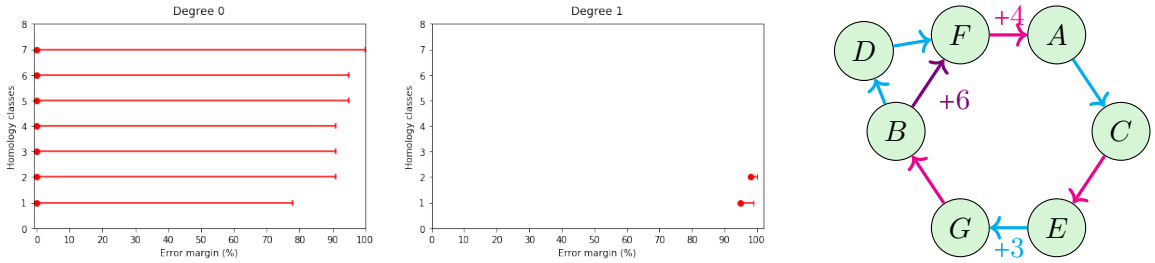


Figure 1: The associated family of barcodes with the C-major scale (left) and the graph of the filtration with a scale of 95% (right).

We then extended this construction to musical bars containing 3-chords from a specific space, such as Euler's Tonnetz, or more generally any two-dimensional Tonnetz of the classical form $T[a, b, c]$. As in the case of the scales, the barcodes here reveal consistent musical features: more precisely, we find back the topological structure associated with the Tonnetz, such as the torus in the case of $T[3, 4, 5]$ shown in Figure 2. In this particular case, we also recover the PLR-group of fundamental transformations on the Tonnetz, as illustrated by Theorem 1.

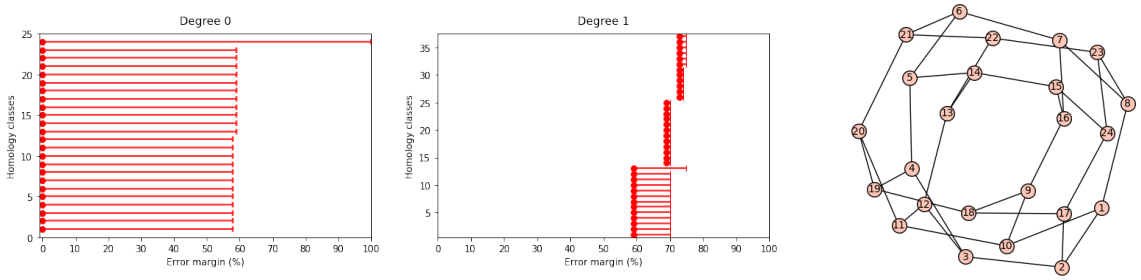


Figure 2: The associated family of barcodes with the Euler's Tonnetz (left) and the graph of the filtration with a scale of 59% (right).

Theorem 1. *The graph associated with the twenty-four major and minor chords of the Euler's Tonnetz, given by the Rips filtration and the DFT-distance with a scaling parameter of* 59%*, corresponds to the Cayley graph of the PLR-group generated by the three transformations P, L and R.*

★ Prospect and future work

This chapter confirms our belief that the distance thus constructed by means of the DFT, together with persistent homology, seems to be a reasonable tool for understanding known musical structures.

However, there are several other artificial musical scores on which we might want to test our model. As a natural starting point, there are different musical modes and scales that need to

be tested, and we could imagine making a general classification of most of the common ones. We also computed the DFT-distance on the twelve two-dimensional Tonnetze, and there are further generalizations to be made here: as we did for the chromatic scale, we could also try the metric on the set of all the 3-chords in $\mathbb{Z}/12\mathbb{Z}$ (the 220 possibilities). For a given 3-chord, the closest would still be its relative, then its leading-tone and its parallel, but this will allow us to classify all the 220 chords with our metric, and this will also give a meta-classification between the twelve Tonnetze. Finally, the DFT-distance could also be applied to n-chords, or simply to higher dimensional Tonnetze, as presented in [53].

To go further, we could also imagine trying our distance on musical objects that focus more on rhythms, such as the general *Zeitnetze* that is introduced in [57].

Harmonization, graph-type and complexity

This chapter presents a new way of analyzing some musical pieces from the specific register of Pop music, and measuring their complexity using associated graph-types built from barcodes. The harmonization of Pop songs, i.e. their reduction to melody and accompaniment, allows us to take several points of study, especially by considering scores with increasing levels of analysis.

★ The results

The first idea was to look at different tracks from a given score: the chord chart, the accompaniment and the complete harmonized song. By comparing these levels of analysis, we can observe what the accompaniment or the melody brings to the song. In the example of Figure 3, the studied song becomes more complex from level 0 to level 1, and the vertices are classified by chord types at level 2. This last observation is common to most of the songs we analyzed.

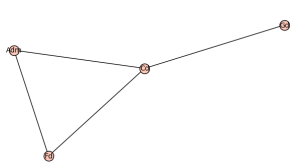
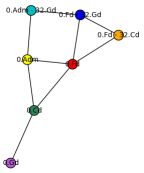
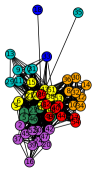
	Chord Chart	**Accompaniment**	**With Melody**
Happy Ending			
	87%	63%	69%

Figure 3: Comparison of the three levels of analysis for the song *Happy Ending*.

The comparison between different levels of analysis gave rise to the notion of complexity: in fact, we set up an algorithm based on barcodes in degrees 0 and 1 that produces a graph-type for each song (for a given level of analysis). By comparing these graphs, we can then distinguish several songs according to their complexity, that means their 1-Betti numbers. For instance, Figure 4 shows different graph-types for four different songs at level 1 of analysis, and Figure 5 illustrates the notion of complexity with histograms representation.

★ Prospect and future work

Associating a graph with a song sounds like a reasonable idea, but we have to be careful here to avoid the approach of providing a "picture" of the song at a fixed moment, and giving up the temporal dimension. Here the construction is different: we first build a filtration, and then use barcodes to choose the most representative moment to look at our song.

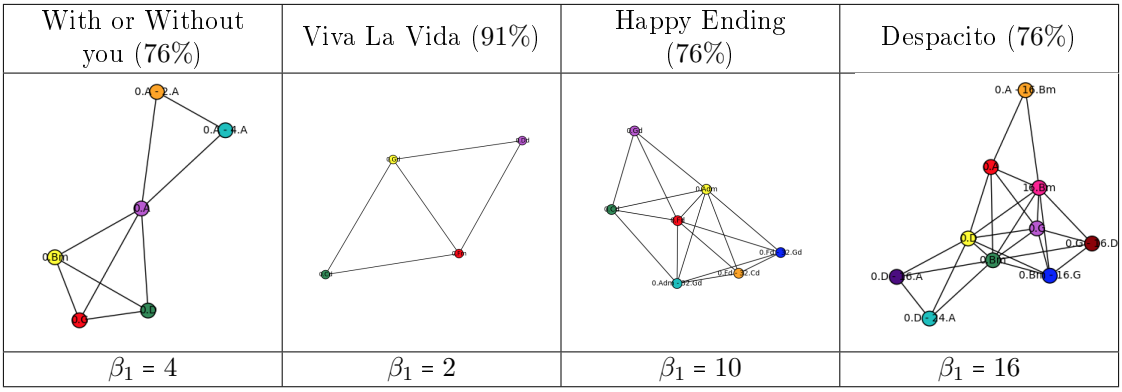


Figure 4: The associated graph-types and Betti numbers β_1 for different songs based on four chords at level 1 of analysis (accompaniment).

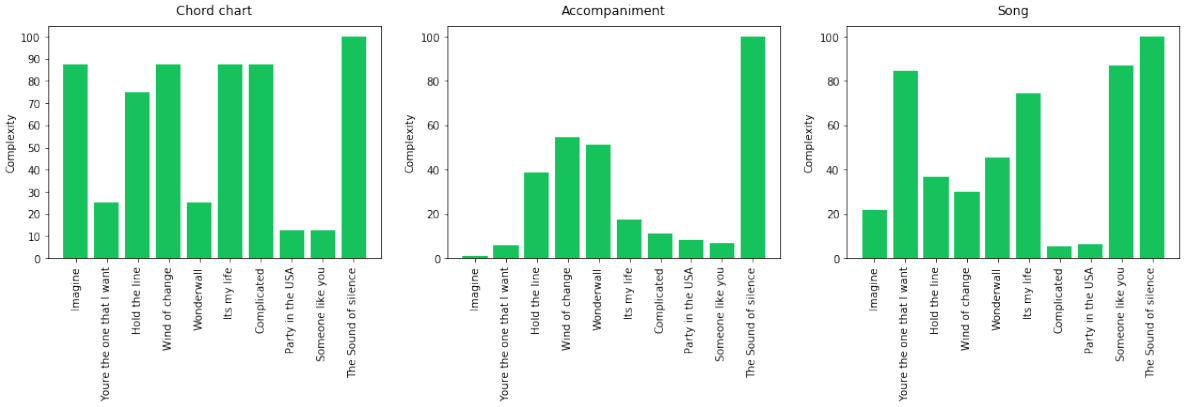


Figure 5: The complexity of ten songs based on six chords for each level of analysis.

In future work, we plan to consider higher dimensions of homology and not limit the study to only degrees 0 and 1: indeed, we can imagine an algorithm that provides a graph-type depending on the degree of homology we choose to stop at. This will necessarily have an impact on our definition of complexity, which we will then have to adjust by looking at higher Betti numbers. Furthermore, if we worked with a small number of chords, we could also apply the harmonization process to songs based on N-chords, for larger values of N. In the same way, we need to consider graphs that include different types of chords (seventh chords, suspended chords, etc.), that is, not restrict the study to chords that come only from Euler's Tonnetz.

Automatic music style classification

This chapter deals with the well-known problem of automatic classification of musical styles. Here we propose to confront our DFT-approach to this problem by converting the resulting barcode (in degree 0) associated with a score into a point in $\mathbb{R}^3$, which is expressed by three basic statistics on the length of the bars (mean value, standard deviation and entropy). This idea allows us to visualize a family of MIDI files in $\mathbb{R}^3$, and then to analyze the distribution of the points to see possible clusters.

★ The results

The proposed method gives satisfactory results when the artists studied come from distant styles, as it is shown in Figure 6, which has the merit of confirming that the metric used seems musically

consistent. When the chosen styles are more difficult to distinguish, such as when comparing Pop/Rock with Heavy Metal, or when the number of MIDI files increases, it becomes more difficult to visually establish clusters. We then set up a method to create clustering trees from several families of scores, by taking into account the distance between the average points for each style studied. The results are illustrated with the classification trees from Figure 7.

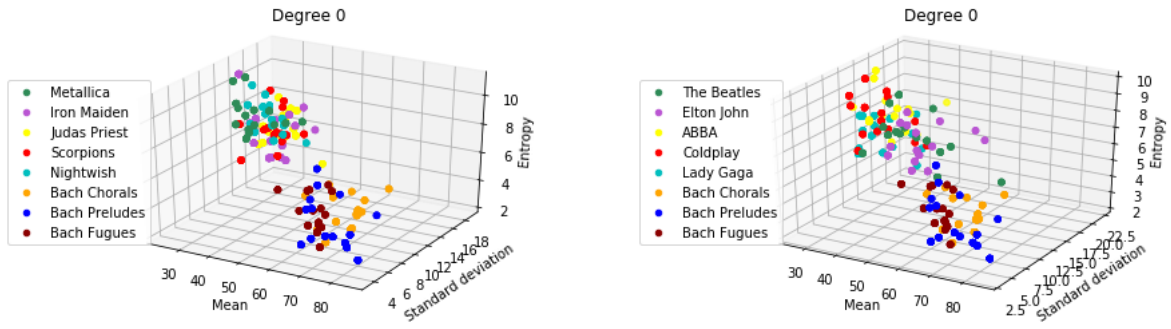


Figure 6: Classification of different musical styles by clustering in $\mathbb{R}^3$: Heavy Metal VS Baroque (left) and Pop VS Baroque (right).

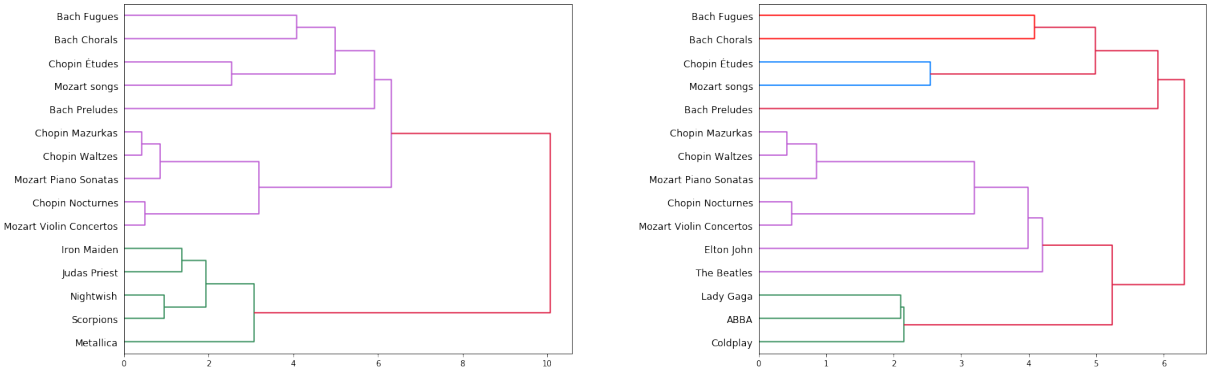


Figure 7: Classification of different musical styles by clustering trees: Heavy Metal VS Classical (left) and Pop VS Classical (right).

★ Prospect and future work

Our approach seems to be musically consistent, considering the classification results. However, we have to keep in mind that the tests were done with a small amount of data for each style (fifteen MIDI files for each artist/group), so we plan to push the analysis further by increasing the number of data. In the same idea, we have started to collect different musical styles in our database, such as jazz or even film and video game music, and obviously it would be a natural idea to extend our analysis to other musical styles.

On the other hand, a line of research that needs to be exploited is the degree of homology we are working with: in fact, as in the previous applications, here we only considered the dimension 0 here, which has the merit to produce convincing results, but perhaps it could be interesting to understand what other degrees might bring to the analysis. This would certainly change our clustering tree and perhaps allow us to obtain even more precise results. Similarly, we could work on the musical interpretation of the mean, the standard deviation and the entropy in the case of dimension 0, but also in higher degrees.

Hausdorff distance and scores global structure

This chapter proposes to extract a filtered complex from a piece of music, using the same musical bar method but with a different metric than the DFT, given by the Hausdorff distance. This new construction shows some interesting results on a precise style of composition, namely minimalist music.

★ The results

By applying the Hausdorff metric to minimalist music, represented in this study by the French composer Yann Tiersen, we found that persistent homology seems to be able to capture the global structure of the piece using barcodes in degree 0. More precisely, if a song is built on a precise model, such as "introduction - theme 1 - theme 2 - theme 3 - conclusion", then the filtration for a well-chosen scale describes this structure. This is illustrated by Figure 8 on which we can read the different themes of the musical piece *Comptine d'un autre été: l'Après-midi*. These results are also shortly presented in [18].

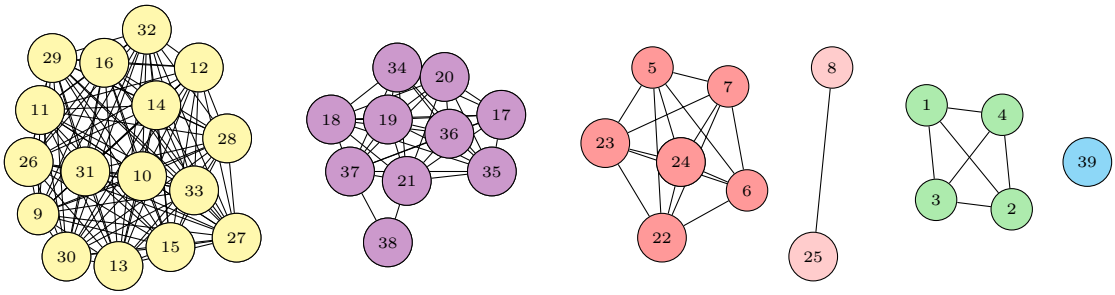


Figure 8: The associated filtration with *Comptine d'un autre été: l'Après-midi*, with a scale of 14%: each component characterizes a theme of the piece.

★ Prospect and future work

The Hausdorff metric provides another type of analysis based on the global structure of a song, which is captured by degree 0. In contrast, barcodes in degree 1 have no obvious musical interpretation yet, and a natural trail for future research would be to focus on this dimension and even higher ones, as for the other musical applications given in this thesis. We believe that one-dimensional cycles could indeed be related to repetitive patterns or musical loops in the score, and we are working to highlight this interpretation from our construction. Furthermore, we could try to relate this work to the notion of musical texture, as presented in the work [22].

On the other hand, this new metric based on the Hausdorff distance seems to be musically consistent in the minimalist musical style, and we could generalize the results to other composers of this style in a first place, but also to a more general and diverse corpus of musical data, as for the classification issue. In this sense, we could try to understand how the Hausdorff distance is able to capture the global structure of stylistically different musical pieces that are not necessarily built on a precise sequence of themes.

Annexe A. Codes

Une partie importante de notre travail a porté sur la mise en œuvre de codes pour nos calculs. Dans cette thèse, nous avons choisi d'utiliser **SageMath**, le système logiciel Sage Mathematics (Version 9.1.0). SageMath est un logiciel libre basé sur le langage Python, qui utilise de nombreux logiciels libres existants tels que NumPy, SciPy, matplotlib et bien d'autres (voir `https://www.sagemath.org` et [51]).

Dans cette annexe, nous ne visons pas à énumérer tous les codes mis en œuvre pour cette thèse, mais simplement à donner un aperçu de ce qui a été construit en présentant les principaux algorithmes. Les codes complets sont disponibles sur la page web dédiée :

`https://math-musique.pages.math.unistra.fr/programmation.html`

MIDI et DFT

La première partie du processus de codage a consisté à la *lecture* d'un fichier MIDI : pour cela, nous avons choisi d'utiliser le package `py-midi` (Version 2.0.1) de Python3, qui permet notamment de recevoir des messages MIDI (voir `https://pypi.org/project/py-midi/`). Plus précisément, étant donné un fichier MIDI en entrée, le module renvoie une liste de triplets de la forme

$$(onset, length, pitch)$$

Cette sortie correspond aux deux premières lignes que nous utilisons dans le code présenté ci-dessous. Avec ce type de données, étant donné un fichier MIDI et un chiffrage correspondant en entrée, nous pouvons écrire un algorithme simple qui extrait les mesures musicales du fichier donné sous la forme d'une liste de paires de sous-listes $[onset, pitch]$ dans $\mathbb{Z}/t\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$, conformément à la définition 1.2.2 du premier chapitre. L'unité de temps t et l'unité de hauteur p peuvent être calculées directement à partir du fichier MIDI en examinant la note la plus courte et l'*ambitus* global de la partition. Les valeurs *onset* et *pitch* sont encodées en utilisant respectivement les notions de *tiks* et *midicent* (voir la section introduction), et peuvent être converties modulo t et p. La fonction finale `FourierBars` est représentée ci-dessous.

```
#a first function for extracting bars for a specific track
import midi

def ExtBarsTrack(filename, track, meter):
    track = MIDI_Track(filename, track) #using midi module
    L = track.notes() #list of tuples (start, end, pitch)
    m = meter[0]  #nbr of beats per bar
    n = meter[1]  #time unit (1=whole,2=half,4=quarter,8=quaver)
    u = 1920/n  #time unit in midi tiks (1920=whole=1)
    d = m*u  #length of a bar in midi tiks
```

```
    #creation of bars_list
    l = len(L)  #nbr of notes in track
    if l==0:
        bars_list = []
    else:
        nbr_bars = floor(L[l-1][1]/int(d)) +1
        bars_list = nbr_bars*[[]]

        #placing each note in the corresponding bar
        for x in L:
            i = int(x[1]/d)
            x.remove(x[1])
            if bars_list[i] == []:
                bars_list[i] = [x]
            else:
                bars_list[i].append(x)

    return bars_list

#the function we use for extracting bars for all tracks
def FourierBars(filename,track_list,meter):
    m = meter[0]
    n = meter[1]
    u = 1920/n
    d = m*u

    onsets = [] #list of onsets
    pitches = [] #list of pitches
    len_t = [] #length of each track t

    #first get the time and pitch-unit
    for t in track_list:
        bars_list_t = ExtBarsTrack(filename,t,meter)
        len_t.append(len(bars_list_t))
        for bar in bars_list_t:
            pitches.append(bar[0][1])
            for i in range(1,len(bar)):
                pitches.append(bar[i][1])
                ons = bar[i][0]-bar[i-1][0]
                if ons!=0:
                    onsets.append(ons)
    if onsets == []:
        time_unit = 0
    else:
        m = min(onsets)
        T = [0,30,60,96,120,160,240,480,960,1920] #the tikz values
        for k in range(len(T)):
            if T[k] < m <= T[k+1]:
                p = T[k+1]
        time_unit = 1920/p

    onset = 1920/time_unit
    pitch_unit = min(pitches) - min(pitches)%12 #minimal pitch C
    len_bars = max(len_t)

    #list of final bars for all tracks
```

```
    Bars = [[] for i in range(len_bars)]
    for t in track_list:
        bars_list_t = dftBarsTrack(filename,t,meter)
        if bars_list_t == []:
            bars_list_t = [[] for i in range(len_bars)]
            for i in range(len(Bars)):
                Bars[i] = Bars[i] + bars_list_t[i]
        else:
            for i in range(len(Bars)):
                if len(bars_list_t) < len_bars:
                    for k in range(1-len(bars_list_t)):
                        brs_list_t.append([])
                    Bars[i] = Bars[i] + bars_list_t[i]
                else:
                    Bars[i] = Bars[i] + bars_list_t[i]

    #transposition modulo time and pitch unit
    for i in range(len(Bars)):
        for j in range(len(Bars[i])):
            Bars[i][j][0] = ((Bars[i][j][0] - d*i)/onset)%time_unit
            Bars[i][j][1] = ((Bars[i][j][1] - pitch_unit))
        Bars[i].sort()

    dist_bars = []
    for B in Bars:
        if B not in dist_bars and B != []:
            dist_bars.append(B)

    return dist_bars
```

Une fois que la liste de toutes les mesures musicales d'un fichier MIDI donné est extraite, nous utilisons le module NumPy `numpy.fft2` pour calculer la DFT bidimensionnelle, de sorte à créer un algorithme qui calcule la DFT-distance de la définition 1.2.5 entre deux mesures musicales données. Le résultat est donné sous forme d'un dictionnaire dont les clés sont les paramètres d'échelle (de 0 à 100 après le processus de normalisation, voir la section 4.3) et, pour chaque clé, la valeur est la paire de mesures musicales correspondante. La fonction finale `FourierMetricNorm` prend en entrée une liste de mesures musicales extraites via la méthode proposée ci-dessus, une unité de temps et une unité de hauteur.

```
import numpy as np

#characteristic function
def carFunctionBar(bar,u_time,u_pitch):
    M = np.zeros((u_time,u_pitch))
    for i,j in bar:
        M[i,j] = 1

    return M

#DFT of a musical bar
def dft(bar,u_time,u_pitch):
    A = carFunctionBar(bar,u_time,u_pitch)
    if u_time ==1:
        dft = np.fft.fft(A)
    else:
        dft = np.fft.fft2(A)
```

```
    return dft

#DFT between two musical bars
def dftMetricBars(bar1,bar2,u_time,u_pitch):
    A = dft(bar1,u_time,u_pitch)
    B = dft(bar2,u_time,u_pitch)
    N = np.abs(A-B)
    M = N.sum()

    return M

#the final function that provides the dictionary of distances
def FourierMetricNorm(bars_list,u_time,u_pitch):
    nbr_bars = len(bars_list)

    #list of triplet [distance(i,j), bar_i, bar_j)
    dist = []
    for i in range(nbr_bars):
        for j in range(i+1,nbr_bars):
            dist.append([dftMetricBars(bars_list[i],bars_list[j],u_time,
                u_pitch))

    epsilon = [e for e in dist]
    M = max(epsilon)
    prec = (M/100) #precision we work with
    dist_prec = dict()
    dist_prec[0] = [(i+1) for i in range(nbr_bars)]  #initialisation

    for i in range(len(bars_list)):
        for j in range(i+1,nbr_bars):
            d = dftMetricBars(bars_list[i],bars_list[j],u_time,u_pitch)
            k = int(d/prec)  #k*prec <= d < (k+1)*prec
            if (k==0) and 0 not in dist_prec:
                dist_prec[1] = []
                dist_prec[1] = [(i+1,j+1)]
            elif (k==0) and 0 in dist_prec:
                dist_prec[1] = []
                dist_prec[1].append((i+1,j+1))
            elif k not in dist_prec:
                dist_prec[k] = [(i+1,j+1)]
            elif k in dist_prec:
                dist_prec[k].append((i+1,j+1))

    #induction
    for k in range(1,101):
        if k not in dist_prec:
            dist_prec[k] = dist_prec[k-1]

    #sorting the dictionnary
    dist_prec_sort = dict()
    L = list(dist_prec.keys())
    L.sort()
    for k in L:
        dist_prec_sort[k]=dist_prec[k]

    return dist_prec_sort
```

Filtration et codes-barres

À partir d'un fichier MIDI donné, nous pouvons maintenant extraire la liste des mesures musicales et calculer la DFT-distance associée. Cette dernière information est stockée dans un dictionnaire, où les clés sont les paramètres d'échelle et les valeurs sont les paires de mesures musicales correspondantes. Ainsi, nous sommes en mesure d'extraire un nuage de points d'un fichier MIDI, et nous donnons ici le code qui nous permet de calculer la filtration de Vietoris-Rips correspondante.

```
def RipsFiltration(bars_list,dft_dict):
    bars = bars_list #list of distinct musical bars, modulo t and p
    dist = dft_dict #dictionnary of DFT-distances with pair of bars
    nbr_bars = len(bars)
    vertices = [i+1 for i in range(nbr_bars)] #a vertex = a musical bar

    epsilon = [e for e in dist] #scaling parameters
    epsilon.remove(0)
    epsilon.sort()

    #initialisation = complex at time 0 with only vertices
    graph_0 = Graph()
    graph_0.add_vertices(vertices)

    filt_graph = [graph_0] #list of graphs
    filt_complex = dict() #list of complexes
    filt_complex[0] = {'vertices':vertices,'edges':[],'triangles':[]}

    for i,e in enumerate(epsilon, start = 1):
        #for each scale e put the corresponding edges to the ith graph
        graph_i = filt_graph[i-1].copy()
        if type(dist[e][0]) == Integer:
            graph_i.add_vertices(dist[e])
        else:
            graph_i.add_edges(dist[e])

        filt_graph.append(graph_i)

        #for each parameter "e" we create the corresponding ith complex
        edges_e=[]  #edges list for the ith complex at scale e
        E = filt_graph[i].edges(labels=False)
        for k in range(len(E)):
            edges_e.append(list(E[k]))

        triangles_e=[] #triangles list for the ith complex at a scale e
        triangles = Set([])

        S = filt_graph[i].vertices()
        gr = filt_graph[i].copy()

        for s in S:
            Ls = gr.neighbors(s)
            Gs = gr.subgraph(Ls)
            Es = Gs.edges(labels = False)

            triangles = triangles.union( Set([ Set([s, edge[0], edge
```

```
                [1]]) for edge in Es ]) )
            gr.delete_vertex(s)

        for t in list(triangles):
            triangles_e.append(list(t))

        filt_complex[e] = {'vertices' : vertices, 'edges' : edges_e, '
            triangles' : triangles_e}

    return filt_complex
```

Il est important de noter que nous n'ajoutons que les simplexes de dimensions 0, 1 et 2 (sommets, arêtes et triangles), puisque nous n'avons besoin de calculer que l'homologie de degré 0 et 1 dans nos applications. Pour afficher les codes-barres correspondants, nous utilisons un module nommé `persil` implémenté par G. Rousseau (voir [45]), qui est basé sur l'article de G. Carlsson et A. Zomorodian ([59]). En effet, ce module nous permet de calculer à partir d'une filtration de Rips la liste des "$\mathscr{P}$-intervalles", qui correspondent simplement à la liste des paires $(b_{i,d}, d_{i,d})$ pour $b_{i,d}$ et $d_{i,d}$ respectivement la naissance et la mort de la ith classe d'homologie de degré d. L'algorithme ci-dessous montre comment nous récupérons ces $\mathscr{P}$-intervalles en utilisant le module `persil`. Les codes-barres sont facilement tracés à partir de ces intervalles à l'aide du package matplotlib.

```
def RipsIntervals(filtration,degree):
    F = filtration
    N = len(F) #number of complexes (generally 100)

    list_simplex_degree = []  #list of pairs (simplex,apparition time)

    #initialisation (vertices at time 0)
    sommets = F[0]['vertices']
    for v in sommets:
        list_simplex_degree.append(([v],0))

    #induction
    for t in range(1,N):
        E_t = F[t]['edges']
        for edge in E_t:
            list_simplex_degree.append((edge,t)) #edges

        T_t = F[t]['triangles']
        for triangle in T_t:
            list_simplex_degree.append((triangle,t)) #triangles

    #creation of barcodes in degree d using persil module
    fc = FilteredComplex()
    for (simplex, value) in list_simplex_degree:
        fc.insert(simplex,value)

    zc = ZomorodianCarlsson(fc,strict = True) #persil module
    zc.computeIntervals()

    return zc.intervals[d]
```

Classification du style musical

Grâce aux codes présentés dans le paragraphe précédent, nous sommes en mesure de calculer les codes-barres d'un fichier MIDI, c'est-à-dire les éléments de la forme

$$\{(b_i, d_i) \mid b_i \text{ naissance de la classe } i,\ d_i \text{ mort de la classe } i\} \subset \mathbb{N}^2.$$

Dans le chapitre 7, nous proposons de calculer plusieurs statistiques sur la longueur des barres des codes-barres (moyenne, écart-type et entropie). Ces statistiques nous permettent de classer une famille de morceaux de musique en les regroupant dans $\mathbb{R}^3$. Le premier algorithme montre comment nous calculons les valeurs moyennes d'une famille de fichiers MIDI sous la forme d'un dictionnaire. Plus précisément, `dict_artist` est un dictionnaire dont les clés sont les titres des chansons (les fichiers MIDI) et les valeurs sont des dictionnaires de la forme

`{'degree0':[barcode_degree0],'degree1':[barcode_dergee1]}`

Bien entendu, nous pouvons écrire des algorithmes similaires pour l'écart-type et l'entropie. Le second algorithme montre comment nous pouvons représenter plusieurs familles de fichiers MIDI dans $\mathbb{R}^3$, en entrant un dictionnaire `adversaires` dont les clés sont les `dict_artiste` tels que définis ci-dessous. Pour les résultats présentés dans cette thèse, nous avons également écrit des algorithmes similaires pour tracer les points dans $\mathbb{R}^2$.

```
def PersMeanBarcodes(dict_artist,degree):
    mean = {}
    for midi in dict_artist:
        m = []
        D = dict_artist[title]['degree0']
        s = 0
        for i in range(len(D))):
            s = s+(D[i][1]-D[i][0])

        m.append(s/len(D))
        mean[title] = [m]

    return mean
```

```
import matploblib.pyplot as plt
import matploblib.colors as mcolors

def PersComparison3D(adversaries,degree):
    fig = plt.figure()
    ax = fig.add_subplot(111, projection='3d')
    colors = list(mcolors.values())

    A = list(adversaries.keys())

    for i in range(len(A)):
        #statistical values for each adversary of the comparison
        mean = []
        dev = []
        ent = []

        for j in range(len(A[i])):
```

```
            a = adversaries[A[i]]
            mean_adversary = PersMeanBarcodes(a)
            dev_adversary = PersDevBarcodes(a)
            ent_adversary = PersEntBarcodes(a)
            for title in mean_adversary:
                mean.append(mean_adversary[title][degree])
                deg.append(dev_adversary[title][degree])
                ent.append(ent_adversary[title][degree])

        ax.scatter(mean, var, ent, c=colors[i], label = A[i])

    ax.set_xlabel('Mean')
    ax.set_ylabel('Standard deviation')
    ax.set_zlabel('Entropy')
    ax.legend()
```

Graph-type et fichiers MIDI

Le dernier code que nous présentons dans cette annexe est celui qui fournit ce que nous avons appelé un « graph-type » associé à n'importe quel fichier MIDI. En effet, dans la section 6.4, nous avons proposé une méthode pour associer un certain graphe à un fichier MIDI donné : ce protocole est présenté dans 6.4.1 et est basé sur l'interprétation persistante des codes-barres, c'est-à-dire la longueur des barres. L'implémentation de cette approche est présentée ci-dessous, et nécessite les fonctions de la première partie (`FourierBars`, `FourierMetricNorm`, `RipsFiltration` et `RipsIntervals`).

```
#list of associated graphs for a filtration
def RipsGraphs(filtration):
    list_graph = []
    nbrComplexes = len(filtration)
    for t in range(nbrComplexes):
        C = filtration[t]
        G = Graph()
        G.add_vertices(C['vertices'])
        G.add_edges(C['edges'])
        list_graph.append(G)

    return list_graph

#the graph-type for a midi file
def MusicGraph(filename,track_list,meter,u_time,u_pitch):
    bars_list = FourierBars(filename,track_list,meter)
    dist = FourierMetricNorm(bars_list,u_time,u_pitch)
    filtration = RipsFiltration(bars_list,dft_dict)
    graph = RipsGraphs(filtration)

    inter_1 = RipsIntervals(filtFilename,1)
    inter_0 = RipsIntervals(filtFilename,0)
    inter_0e = [inter_0[i][1] for i in range(len(inter_0))] #end of each
        bar in degree 0
    inter_1s = [inter_1[i][0] for i in range(len(inter_1))] #start of
       each bar in degree 1

    #longest bar for H0
```

```
    m0 = {}
    for i in range(1,len(inter_0e)):
        time = inter_0e[i]-inter_0e[i-1]
        if time not in m0:
            m0[time] = [inter_0e[i-1]]
        else:
            m0[time].append(inter_0e[i-1])
    m00 = max(list(m0.keys()))
    t0 = m0[m00][0]

    #truncation of H1 from t0
    L = [i for i in range(101)]
    L0 = L[t0:]
    m = []
    for i in L0:
        if i in inter_1s:
            m.append(i)
    s = inter_1s.index(min(m))
    inter_1s = inter_1s[s:]

    #longest bar for H1
    m1 = {}
    for i in range(1,len(inter_1s)):
        time = inter_1s[i]-inter_1s[i-1]
        if time not in m1:
            m1[time] = [inter_1s[i-1]]
        else:
            m1[time].append(inter_1s[i-1])

    m11 = max(list(m1.keys()))
    t1 = m1[m11][0]
    if t1 == 100:
        t1 = t0

    #final graph-type
    print('Error margin t0 (%):', t0)
    print('Error margin t1 (%):', t1)

    return graph[t1]
```

Annexe B. Base de données

Dans le cadre des travaux de cette thèse, nous avons été confrontés au problème des bases de données de fichiers MIDI : en effet, pour effectuer des tests sur la distance et le complexe simplicial filtré que nous construisons, il est nécessaire de tester nos programmes sur un très grand nombre de fichiers. Il existe cependant très peu de bases de données disponibles en ligne, et encore moins avec une grande variété de styles musicaux. Nous avons donc décidé de consacrer une partie de cette thèse à la construction d'une telle base de données et disponible en ligne. Ce travail tente de répondre à la question de savoir comment collecter des données et les rassembler dans une base de données unique, et il s'agit bien sûr d'un travail qui doit être poursuivi. Nous avons donc créé un site web dédié à cette base de données, accessible à l'adresse suivante :

`https://math-musique.pages.math.unistra.fr/`.

Dans la section "Base de données", nous proposons douze styles musicaux différents : Classique, Electro, Folk/Country, Hard Rock/Heavy Metal, Jazz/Blues/Soul, Musique de film et de série, Musique de jeu vidéo, Pop/Rock, Rap, Reggae, RnB/Funk/Disco et Variété française. Cette classification des fichiers MIDI est illustrée via la figure B.1. Par exemple, la figure B.2 montre les fichiers MIDI disponibles pour les styles Hard Rock et Heavy Metal, utilisés notamment dans le cadre de la classification automatique du chapitre 7.

Figure B.1: Le menu de la base de données des fichiers MIDI avec douze styles musicaux différents (`https://math-musique.pages.math.unistra.fr/midi.html`).

La base de données proposée se compose de fichiers MIDI collectés sur différents sites tels que *Musescore*, *Kunstderfuge* ou *Bitmidi*. Pour chaque morceau, le groupe ou l'artiste est indiqué, ainsi que l'année de composition, le style et la source.

A noter que dans le cadre de notre travail, nous avons fait le choix de supprimer les pistes de batterie et de percussion de nos fichiers MIDI, et ce dans le but de travailler avec une représentation symbolique. Pour retrouver le fichier original et complet, il suffit de se rendre sur le site source indiqué dans la bibliographie.

Figure B.2: Un exemple de fichiers MIDI disponibles pour les styles Hard Rock et Heavy Metal (`https://math-musique.pages.math.unistra.fr/midi/hardrock-metal-midi.html`).

Bibliography

[1] Mehmet Emin Aktas, Esra Akbas, and Ahmed El Fatmaoui. Persistence homology of networks: Methods and applications. https://doi.org/10.48550/arXiv.1907.08708, 2019.

[2] Alberto Alcalá-Alvarez and Pablo Padilla-Longoria. A framework for topological music analysis (TMA). https://doi.org/10.48550/arXiv.2204.09744, 2022.

[3] Emmanuel Amiot. Rhythmic canons and Galois theory. In *Colloquium on Mathematical Music Theory*, volume 347 of *Grazer Math. Ber.*, pages 1–33. Karl-Franzens-Univ. Graz, Graz, 2005.

[4] Emmanuel Amiot. *Music through Fourier space*. Computational Music Science. Springer, Cham, 2016.

[5] Emmanuel Amiot. Entropy of Fourier coefficients of periodic musical objects. *Journal of Mathematics & Music. Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance*, 15(3):235–246, 2021.

[6] Milton Babbitt. Some aspects of twelve-tone composition. *The Score and I.M.A. Magazine*, (12):53–61, 06 1955. London: Score Publishing, International Music Association.

[7] Robin Lynne Belton, Brittany Terese Fasy, Rostik Mertz, Samuel Micka, David L. Millman, Daniel Salinas, Anna Schenfisch, Jordan Schupbach, and Lucia Williams. Learning simplicial complexes from persistence diagrams. https://doi.org/10.48550/arXiv.1805.10716, 2018.

[8] Mattia Bergomi. *Dynamical and Topological Tools for (Modern) Music Analysis*. PhD thesis, Università degli Studi di Milano; Université Pierre et Marie Curie, December 2015. Computer Science.

[9] Mattia G. Bergomi and Adriano Baratè. Homological persistence in time series: an application to music classification. *Journal of Mathematics & Music. Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance*, 14(2):204–221, 2020.

[10] Mattia G. Bergomi, Adriano Baratè, and Barbara Di Fabio. Towards a topological fingerprint of music. In *Computational topology in image context*, volume 9667 of *Lecture Notes in Comput. Sci.*, pages 88–100. Springer, [Cham], 2016.

[11] Louis Bigo. *Représentations symboliques musicales et calcul spatial*. PhD thesis, Université Paris-Est, 12 2013.

[12] Louis Bigo and Moreno Andreatta. A Geometrical Model for the Analysis of Pop Music . *Sonus*, 35(1):36–48, 2014.

[13] Louis Bigo and Moreno Andreatta. Topological structures in computer-aided music analysis. In David Meredith, editor, *Computational music analysis*, pages 57–80. Springer, Cham, 2016.

[14] Louis Bigo and Moreno Andreatta. Filtration of pitch-class sets complexes. In Mariana Montiel, Francisco Gomez-Martin, and Octavio A. Agustín-Aquino, editors, *Mathematics and computation in music*, volume 11502 of *Lecture Notes in Comput. Sci.*, pages 213–226. Springer, Cham, 2019.

[15] Louis Bigo, Moreno Andreatta, Jean-Louis Giavitto, Olivier Michel, and Antoine Spicher. Computation and visualization of musical structures in chord-based simplicial complexes. In Jason Yust, Jonathan Wild, and John Ashley Burgoyne, editors, *Mathematics and computation in music*, volume 7937 of *Lecture Notes in Comput. Sci.*, pages 38–51. Springer, Heidelberg, 2013.

[16] Louis Bigo, Daniele Ghisi, Antoine Spicher, and Moreno Andreatta. Representation of musical structures and processes in simplicial chord spaces. *Computer Music Journal*, 39(3):9–24, 2015.

[17] Louis Bigo, Jean-Louis Giavitto, and Antoine Spicher. Building topological spaces for musical objects. In Carlos Agon, Moreno Andreatta, Gérard Assayag, Emmanuel Amiot, Jean Bresson, and John Mandereau, editors, *Mathematics and computation in music*, volume 6726 of *Lecture Notes in Comput. Sci.*, pages 13–28. Springer, Heidelberg, 2011.

[18] Victoria Callet. Persistent homology on musical bars. In Mariana Montiel, Octavio A. Agustín-Aquino, Francisco Gómez, Jeremy Kastine, Emilio Lluis-Puebla, and Brent Milam, editors, *Mathematics and Computation in Music*, pages 349–355, Cham, 2022. Springer International Publishing.

[19] Yueqi Cao, Anthea Monod, Athanasios Vlontzos, Luca Schmidtke, and Bernhard Kainz. Topological information retrieval with dilation-invariant bottleneck comparative measures. *Information and Inference: A Journal of the IMA*, 12(3), apr 2023.

[20] Gunnar Carlsson. Topological pattern recognition for point cloud data. *Acta Numerica*, pages 289—368, 2014.

[21] Michael J. Catanzaro. Generalized *tonnetze*. *Journal of Mathematics & Music. Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance*, 5(2):117–139, 2011.

[22] Louis Couturier, Louis Bigo, and Florence Levé. A dataset of symbolic texture annotations in Mozart piano sonatas. In *International Society for Music Information Retrieval Conference*, 2022.

[23] Herbert Edelsbrunner and John Harer. Persistent homology—a survey. In *Surveys on discrete and computational geometry*, volume 453 of *Contemp. Math.*, pages 257–282. Amer. Math. Soc., Providence, RI, 2008.

[24] Herbert Edelsbrunner and John L. Harer. *Computational topology*. American Mathematical Society, Providence, RI, 2010. An introduction.

[25] Herbert Edelsbrunner and D. Morozov. Persistent homology: theory and practice. *European Congress of Mathematics*, pages 31–50, 01 2014.

[26] Massimo Ferri. Persistent topology for natural data analysis — a survey. In Andreas Holzinger, Randy Goebel, Massimo Ferri, and Vasile Palade, editors, *Towards Integrative Machine Learning and Knowledge Extraction*, pages 117–133, Cham, 2017. Springer International Publishing.

[27] Robert Ghrist. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc. (N.S.)*, 45(1):61–75, 2008.

[28] Pierre Guillot. *Leçons sur l'homologie et le groupe fondamental*, volume 29 of *Cours Spécialisés [Specialized Courses]*. Société Mathématique de France, Paris, 2022.

[29] Julien Junod, Pierre Audétat, Carlos Agon, and Moreno Andreatta. A generalisation of diatonicism and the discrete fourier transform as a mean for classifying and characterising musical scales. In Elaine Chew, Adrian Childs, and Ching-Hua Chuan, editors, *Mathematics and Computation in Music*, pages 166–179, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[30] Michael Kerber, Dmitriy Morozov, and Arnur Nigmetov. Geometry helps to compare persistence diagrams. *ACM J. Exp. Algorithmics*, 22:Art. 1.4, 20, 2017.

[31] Koen Klaren. Computational topology in music analysis. https://api.semanticscholar.org/CorpusID:221711387, 2018. Report for Seminar Algorithms (21MA00).

[32] Kai T. Kono, Tsutomu T. Takeuchi, Suchetha Cooray, Atsushi J. Nishizawa, and Koya Murakami. A study on the baryon acoustic oscillation with topological data analysis. https://doi.org/10.48550/arXiv.2006.02905, 2020.

[33] Paul Lascabettes. Homologie persistante appliquées à la reconnaissance de genres musicaux. Master's thesis, École normale supérieure Paris-Saclay, 2018.

[34] Ann B. Lee, Kim Steenstrup Pedersen, and David Mumford. The nonlinear statistics of high-contrast patches in natural images. *International Journal of Computer Vision*, 54(1), 2001.

[35] David Lewin. Re: Intervallic relations between two collections of notes. *Journal of Music Theory*, 3:298, 1959.

[36] David Lewin. Special cases of the interval function between pitch-class sets x and y. *Journal of Music Theory*, 45:1–29, 2001.

[37] John Mandereau, Daniele Ghisi, Emmanuel Amiot, Moreno Andreatta, and Carlos Agon. Discrete phase retrieval in musical structures. *Journal of Mathematics & Music. Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance*, 5(2):99–116, 2011.

[38] John Mandereau, Daniele Ghisi, Emmanuel Amiot, Moreno Andreatta, and Carlos Agon. *Z*-relation and homometry in musical distributions. *Journal of Mathematics & Music. Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance*, 5(2):83–98, 2011.

[39] Martín Mijangos, Alessandro Bravetti, and Pablo Padilla. Musical stylistic analysis: A study of intervallic transition graphs via persistent homology. https://doi.org/10.48550/arXiv.2204.11139, 2022.

[40] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, (1), 2017.

[41] Athanase Papadopoulos and Xavier Hascher. *Leonhard Euler, mathématicien, physicien et théoricien de la musique*. CNRS, 2015.

[42] Alexandre Popoff, Moreno Andreatta, and Andrée Ehresmann. Relational poly-Klumpenhouwer networks for transformational and voice-leading analysis. *Journal of Mathematics & Music. Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance*, 12(1):35–55, 2018.

[43] Ian Quinn. General equal-tempered harmony (introduction and part i). *Perspectives of New Music*, 44(2):114–158, 2006.

[44] Bastian Rieck and Heike Leitte. Enhancing comparative model analysis using persistent homology. In *Proceedings of Workshop on Visualization for Predictive Analytics at IEEE VIS*, 11 2014.

[45] Guillaume Rousseau. `https://github.com/Quenouillaume/persil`, 2021.

[46] X. Serra, M. Magas, E. Benetos, M. Chudy, S. Dixon, A. Flexer, E. Gomez, F. Gouyon, P. Herrera, S. Jorda, O. Paytuvi, G. Peeters, J. Schlüter, H. Vinet, and G. Widmer. *Roadmap for Music Information ReSearch*. MIRES Consortium, London, UK, 2013.

[47] William A. Sethares and Ryan Budney. Topology of musical data. *Journal of Mathematics & Music. Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance*, 8(1):73–92, 2014.

[48] Gurjeet Singh, Facundo Mémoli, Tigran Ishkhanov, Guillermo Sapiro, Gunnar Carlsson, and Dario Ringach. Topological analysis of population activity in visual cortex. *Journal of vision*, 8:11.1–18, 02 2008.

[49] Oleg Starostenko and Omar Lopez-Rincon. A 3d spatial visualization of measures in music compositions. In *Technology, Science, and Culture: A Global Vision*, page 127. IntechOpen, 03 2019.

[50] Audrey Terras. *Fourier analysis on finite groups and applications*, volume 43 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1999.

[51] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 9.1.0)*. `https://www.sagemath.org`.

[52] Mai Lan Tran, Changbom Park, and Jae-Hun Jung. Topological data analysis of korean music in jeongganbo: a cycle structure, 2021.

[53] Dmitri Tymoczko. The generalized tonnetz. *Journal of Music Theory*, 56:1–52, 06 2012.

[54] Dmitri Tymoczko and Jason Yust. Fourier phase and pitch-class sum. In Mariana Montiel, Francisco Gomez-Martin, and Octavio A. Agustín-Aquino, editors, *Mathematics and Computation in Music: 7th International Conference, MCM 2019, Madrid, Spain, June 18–21, 2019, Proceedings*, pages 46––58. Springer-Verlag, Berlin, Heidelberg, 2019.

[55] Dan Tudor Vuza. Supplementary sets and regular complementary unending canons (part two). *Perspectives of New Music*, 29:22–49, 01 1991.

[56] Jason Yust. Fourier methods for computational analysis of enharmonicism and orther harmonic properties. In *Journées d'Informatique Musicale, JIM2020*, Actes des Journées d'Informatique Musicale, Strasbourg, France, 10 2020.

[57] Jason Yust. Generalized *tonnetze* and *zeitnetze*, and the topology of music concepts. *Journal of Mathematics & Music. Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance*, 14(2):170–203, 2020.

[58] Afra Zomorodian. Fast construction of the vietoris-rips complex. *Computers & Graphics*, 34(3):263–271, 2010. Shape Modelling International (SMI) Conference 2010.

[59] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.

Université de Strasbourg

Victoria Callet

Modélisation topologique des structures et processus musicaux

École doctorale Mathématiques, sciences de l'information et de l'ingénieur | ED 269 Université de Strasbourg

Résumé

Cette thèse propose des contributions dans différents problèmes issus du domaine de l'analyse topologique de données musicales : les objets étudiés sont des partitions représentées symboliquement par des fichiers MIDI, et les outils utilisés sont la Transformée de Fourier Discrète et l'homologie persistante. Le manuscrit se découpe en trois parties : les deux premières sont consacrées à l'étude des objets mathématiques précédemment cités et à la mise en place du modèle. Plus précisément, la notion de DFT introduite par Lewin est généralisée au cas de la dimension deux, en explicitant le passage d'une mesure d'un morceau à un sous-ensemble de Z/tZ x Z/pZ, ce qui conduit naturellement à une notion de métrique sur l'ensemble des mesures via leurs coefficients de Fourier. De cette construction naît un nuage de points auquel est associé le complexe filtré de Vietoris-Rips, et par suite une famille de codes-barres donnés par l'homologie persistante. Cette approche permet également de généraliser des résultats classiques tels que le lemme de Lewin ou encore le théorème de l'hexachorde de Babitt. La dernière partie de cette thèse est consacrée aux applications musicales du modèle basé sur la DFT : la première expérience consiste à extraire des codes-barres provenant de partitions artificiellement construites, telles que des gammes ou des accords plaqués. Cette étude mène en particulier à l'harmonisation de chansons que l'on réduit à leur mélodie et leur grille d'accords, ce qui permet entre autre de définir les notions de graphe et de complexité d'un morceau. L'homologie persistante se prête également au problème de la classification automatique du style musical, qui sera traité ici sous le prisme de descripteurs symboliques donnés par des statistiques calculées directement sur les codes-barres. Enfin, la dernière application propose un encodage des mesures basé sur la distance de Hausdorff conduisant à l'étude des textures musicales.

Résumé en anglais

This thesis proposes contributions to various problems in the field of topological analysis of musical data: the objects studied are scores represented symbolically by MIDI files, and the tools used are the discrete Fourier transform and persistent homology. The manuscript is divided into three parts: the first two are devoted to the study of the aforementioned mathematical objects and the implementation of the model. More precisely, the notion of DFT introduced by Lewin is generalized to the case of dimension two, by making explicit the passage of a musical bar from a piece to a subset of Z/tZ x Z/pZ which leads naturally to a notion of metric on the set of musical bars by their Fourier coefficients. This construction gives rise to a point cloud, to which the filtered Vietoris-Rips complex is associated, and consequently a family of barcodes given by persistent homology. This approach also makes it possible to generalize classical results such as Lewin's lemma and Babitt's Hexachord theorem. The last part of this thesis is devoted to musical applications of the model: the first experiment consists in extracting barcodes from artificially constructed scores, such as scales or chords. This study leads naturally to song harmonization process, which reduces a song to its melody and chord grid, thus defining the notions of graph and complexity of a piece. Persistent homology also lends itself to the problem of automatic classification of musical style, which will be treated here under the prism of symbolic descriptors given by statistics calculated directly on barcodes. Finally, the last application proposes a encoding of musical bars based on the Hausdorff distance, which leads to the study of musical textures.

Mots-clés : homologie persistante, analyse topologique de données, complexes simpliciaux, filtration, Vietoris-Rips, analyse musicale, classification, transformée de Fourier discrète