INSTITUT DE RECHERCHE ET COORDINATION
ACOUSTIQUE MUSIQUE (IRCAM)

MASTER'S INTERNSHIP REPORT

Mathematical Morphology Applied to Music

Author:
Paul LASCABETTES

Supervisor:
Carlos AGON
Isabelle BLOCH
Moreno ANDREATTA

in the research team of

Musical Representations

February 2019 - July 2019

Acknowledgements

My first acknowledgments go to Carlos Agon and Isabelle Bloch, who were the main supervisors of this internship. I am particularly thankful for their goodness and availability at all times.

Thanks to monthly meetings, I was able to be supervised by Moreno Andreatta, Louis Bigo and Corentin Guichaoua. With many interns added to the group, we were able to get together in a small team focusing on the application of applied mathematics to music. It is in the atmosphere of that group that new ideas to analyse and understand music thanks to morphological tools emerged. That is why every member of this team played a part in the development of the ideas to be found in this report.

I could not forget to mention the fellow students of my masters degree ATIAM, with whom I spent incredible moments and enriching discussions. Thanks also to the Music Representations team for sharing their knowledge by staying in a good mood. I will not forget other interns and PhD students in Ircam with whom I have shared many coffees and lunches. I will also remark on the good vibes in the office. I also thank the numerous Ircam researchers for allowing me to attend many conferences and discussions.

Abstract

Mathematical morphology is a branch of mathematics that is developing mostly in response to problems related to image processing. Being first applied to binary images and then generalized to grayscale images and to other mathematical frameworks, this theory provides methods for filtering, segmentation and pattern recognition. Due to its relatively recent development, there are no direct applications to music at this time. The objective of this report is to suggest axes of application from mathematical morphology to music. In a first place, by representing the music as a piano roll image, it will be possible to directly apply the morphological operators. This makes it possible to isolate the melody, detect certain patterns or find the key. In a second step, a new method of musical analysis, based on variations in musical pitch, will be developed to represent melodic patterns. Using this approach, it is possible to obtain the self-similarity matrix based on pitch variations. Morphological operators are used to detect the main blocks of this matrix. By changing the size of the filters, this makes it possible to obtain different degrees of filtration to get information on the structure of the piece at different scales.

Keywords : Mathematical morphology · Music visualisation · Musical Structure · Musical pattern detection · Musical contour · Self-similarity matrix

Résumé

La morphologie mathématique est un domaine des mathématiques qui se développe en particulier grâce aux problèmes liés aux traitements d'images. D'abord appliquée aux images binaires puis généralisée aux images à niveau de gris et à d'autres structures mathématiques, cette théorie fournit des méthodes de filtrage, de segmentation ou encore de reconnaissance de formes. De par son développement assez récent, il n'existe pas encore d'applications directes à la musique. Ce rapport a pour but de proposer des axes d'applications de la morphologie mathématique à la musique. Tout d'abord, en représentant la musique sous forme d'une image de type piano roll, il sera possible d'appliquer directement les opérateurs morphologiques. Cela permet d'isoler la mélodie, de détecter certains motifs ou encore de trouver la tonalité. Dans un second temps, une nouvelle méthode d'analyse musicale, basée sur les variations du pitch musical, sera développée pour représenter les motifs mélodiques. Grâce à cette approche, il est possible d'obtenir la matrice d'autosimilarité basée sur les variations du pitch. Les opérateurs morphologiques permettent de détecter les principaux blocs de cette matrice. En changeant la taille des filtres, cela permet d'obtenir différents degrés de filtrage pour avoir une information sur la structure du morceau à différentes échelles.

Mots-clés : Morphologie Mathématique · Visualisation musicale · Structure musicale · Détection de motifs musicaux · Contour musical · Matrice d'autosimilarité

List of Symbols

X^c	complement	$X^c = \{x \in E \mid x \notin X\}$
X_t	translate	$X_t = \{x + t \mid x \in X\}$
$\check{X}$	symmetrical	$\check{X} = \{-x \mid x \in X\}$
δ_S	dilation	$\delta_S(X) = X \oplus S = \{x + s \mid x \in X, s \in S\}$
ε_S	erosion	$\epsilon_S(X) = X \ominus S = (X^c \oplus \check{S})^c$
γ_S	opening	$\gamma_S(X) = X \circ S = \delta_S \circ \epsilon_S(X)$
ϕ_S	closing	$\phi_S(X) = X \bullet S = \epsilon_S \circ \delta_S(X)$
η_S	hit-or-miss	$\eta_S(X) = X \otimes S = (X \ominus S) \cap (X^c \ominus S^c)$
τ_S	top-hat	$\tau_S(X) = X - \gamma_S(X)$
$\gamma_{S_1,\ldots,S_n}$	opening generalized	$\gamma_{S_1,\ldots,S_n}(X) = (X \circ S_1) \cup \ldots \cup (X \circ S_n)$

Contents

Introduction

After having spent the first half of the year between Ircam and Télécom Paris as part of the ATIAM master's program, I wanted to continue in this direction for my research internship. As a result of my background as a mathematician at the ENS, I worked on mathematical morphology in the Musical Representation team at Ircam supervised by Carlos Agon and Isabelle Bloch for six months. It is within this team that new ideas emerged to analyze and understand music with morphological tools. This internship report aims to synthesize these different ideas and present new methods of musical analysis related to mathematical morphology. Indeed, morphology, of which Isabelle Bloch is a specialist, being a field of mathematics derived from image processing that is in the process of development, it has not yet had any direct application to music. Therefore, the main objective of this internship was to understand how to apply morphological tools to music and what were the major interests. It was therefore a continuous process of exploration with no concrete purpose and no real previous work to build on.

To begin with, I studied the literature on mathematical morphology to properly understand this theory. The knowledge that I consider essential is summarized and illustrated in Chapter 1. Since morphology applies to images, my first idea, which is the most intuitive, was to represent music as an image and then to apply morphological operations directly to it. That is why in Chapter 2, I worked on piano rolls which are a representation of music in the form of an image. I then noticed that thanks to morphological operations, it is possible to isolate the melody from the chords, to detect a particular melodic pattern or to extract the key. Results are satisfying but it is not necessary to use morphology to obtain this. Thus, the second chapter aims to propose a first approach to musical analysis with morphological tools.

After having achieved the first objective, I then decided to develop a new technique of musical analysis in order to identify melodic patterns. This technique is based on the variation of the pitch during a song. The innovation proposed in Chapter 3 is the generalization of this technique by moving from a melody to a chord sequence where the number of notes can change. For this purpose, I proposed to work with matrices instead of numbers. Thanks to this generalization, melodic patterns become richer and more complex and make it possible to characterize a musical theme. By working on the self-similarity matrix, based on the variation of the pitch, it is possible to detect the different melodic patterns during a piece. Thus, Chapter 4 explains how to use morphological filters to extract information from this matrix. By gradually varying the filter size, it is possible to determine the structure of the song and identify the different melodic patterns that will represent each theme.

To sum up, the first chapter introduces mathematical morphology, the second presents a first direct application while the last two start from an original personal idea developed throughout the internship in order to find melodic patterns.

Nota Bene : A YouTube channel illustrates the concepts discussed in this report. This will allow you to listen to the different pieces of music that are present and to listen to the transformations made with the morphological tools. The link is available by clicking here.

Chapter 1

Introduction to Mathematical Morphology

1.1 Mathematical Morphology on Sets

1.1.1 Erosion and Dilation

The morphological operations result from two basic operators: *dilation* and *erosion*. Based on these two operations, it is possible to generate several other morphological operations using compositions, unions, intersections, or complement of sets. Only the useful notions are recalled here, and more details can be found in [1, 2, 3, 4, 5, 6]. Let $E = \mathbb{R}^n$ or $\mathbb{Z}^n$ and $X \subseteq E$, before defining dilation and erosion, let us start by reintroducing some notions:

- the *complement* of X is: $X^c = \{x \in E \mid x \notin X\}$
- the *translate* of X by $t \in E$ is: $X_t = \{x + t \mid x \in X\}$
- the *symmetrical* of X is: $\check{X} = \{-x \mid x \in X\}$

Let $S \in \mathcal{P}(E)$ be a *structuring element*, the *dilation* δ_S and the *erosion* ε_S by S are respectively defined as:

$$\delta_S : \begin{array}{ccc} \mathcal{P}(E) & \longrightarrow & \mathcal{P}(E) \\ X & \longmapsto & X \oplus S \end{array} \qquad \varepsilon_S : \begin{array}{ccc} \mathcal{P}(E) & \longrightarrow & \mathcal{P}(E) \\ X & \longmapsto & X \ominus S \end{array}$$

where the *Minkowski addition* $\oplus$ [7] and *substraction* $\ominus$ [8] are defined $\forall X, S \in \mathcal{P}(E)$:

$$X \oplus S = \bigcup_{s \in S} X_s = \bigcup_{x \in X} S_x = \{x + s \mid x \in X, s \in S\} = \{x \in E \mid \check{S}_x \cap X \neq \varnothing\}$$

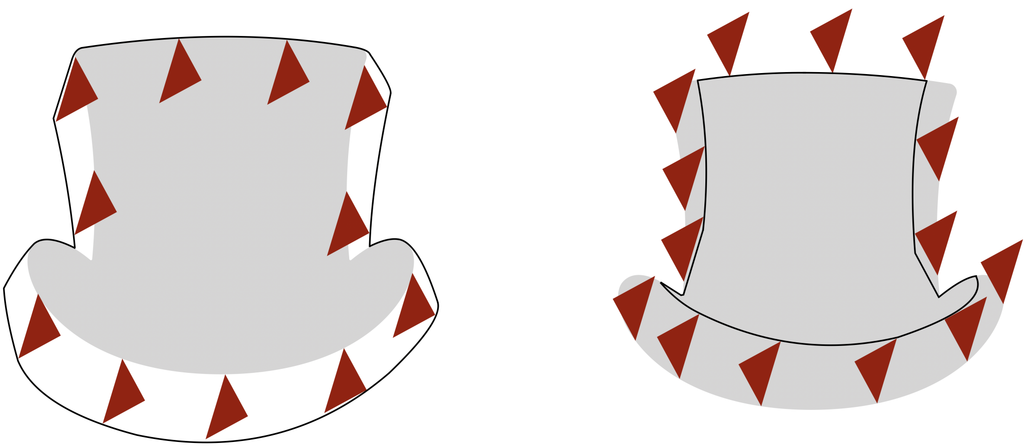
$$X \ominus S = \bigcap_{s \in S} X_{-s} = \{t \in E | S_t \subseteq X\}$$

To illustrate these concepts let X be represented by a hat, and the structuring element S by a triangle (Figure 1.1). It is important to note that the structuring element is defined with an *origin*. Indeed, the position of the origin will have a direct impact on the Minkowski addition and will change the result of a dilation or an erosion. By default, the origin of the structuring element is located at the center of it, however, in the example chosen in Figure 1.1, the origin is located in the upper corner of the structuring element. In addition, if and only if the origin is contained in the structuring element, dilation extends the shapes of the set X, i.e. dilation is *extensive*, while erosion reduces the form of X, erosion is *anti-extensive*. There is a link between these two operations, they are *dual by complementation*, that is to say:

$$X \oplus S = (X^c \ominus \check{S})^c \qquad X \ominus S = (X^c \oplus \check{S})^c$$

FIGURE 1.1: Example of a set X and a structuring element S.

To have an intuition of the shape obtained during a dilation when the origin is contained in the structuring element, one must imagine that the origin of the triangle will slide along the X border. The dilation will be the set of points where the triangle can go. Therefore, the form can only expand. Figure 1.2a illustrates the transformation made by a dilation where the boundary of the dilated set is in black color. In the same way, to visualize erosion, it is possible to use the duality by complementation. Thus, it is necessary to imagine making a dilation of the complementary of X by the symmetrical of S. The origin still slides along the X border but the structuring element is reversed and the boundary of the eroded set will be where the triangle cannot go. The boundary of the erosion result is shown in black in Figure 1.2b. In both cases, the shape has changed and transformations are not just homotheties. To summarize the consequences of a dilation, it fills holes smaller than the



(A) Dilation of X by S

(B) Erosion of X by S

FIGURE 1.2: Example of a dilation and an erosion of X by S.

structuring element, welds nearby shapes, fills narrow channels and widens shapes. On the other hand, for an erosion: it eliminates connected components smaller than the structuring element, widens holes and reduces the size of objects. Finally, if (and only if) the origin is contained in the structuring element:

$$\forall X, S \in \mathcal{P}(E), \quad \varepsilon_S(X) \subseteq X \subseteq \delta_S(X)$$

Dilation and erosion are non-reversible operations. Moreover, they are not reverse but dual operations. Thus, it is possible to represent these links between the two main operators on a graph as shown in figure 1.3.

$$\begin{tikzcd}
\boxed{X} \arrow[r, "\text{dilation by } S"] \arrow[d, dashed, leftrightarrow, "\text{complement}"] & \boxed{\delta_S(X) = X \oplus S} \arrow[d, dashed, leftrightarrow, "\text{complement}"] \\
\boxed{X^c} \arrow[r, "\text{erosion by } \check{S}"'] & \boxed{\varepsilon_{\check{S}}(X^c) = X^c \ominus \check{S}}
\end{tikzcd}$$

FIGURE 1.3: Relation between dilation and erosion on a graph.

1.1.2 Opening and Closing

The other two fundamental operations result from composing the previous functions. In fact, the *opening* γ_S is the composition of an erosion and a dilation. Furthermore, a *closing* ϕ_S is a dilation followed by an erosion.

$$\begin{array}{rccc} \gamma_S: & \mathcal{P}(E) & \longrightarrow & \mathcal{P}(E) \\ & X & \longmapsto & X \circ S \end{array} \qquad \begin{array}{rccc} \phi_S: & \mathcal{P}(E) & \longrightarrow & \mathcal{P}(E) \\ & X & \longmapsto & X \bullet S \end{array}$$

where for all $X, S \in \mathcal{P}(E)$:

$$X \circ S = (X \ominus S) \oplus S \qquad X \bullet S = (X \oplus S) \ominus S$$

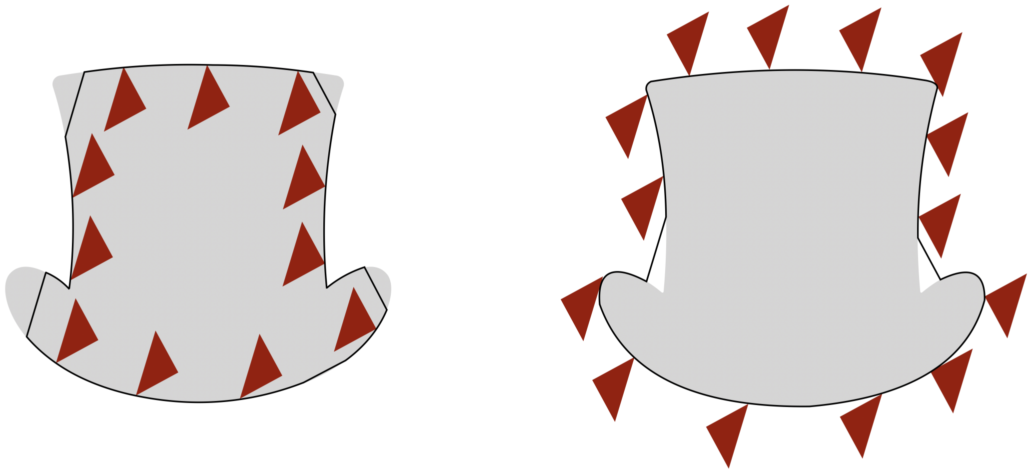
It is easy to interpret the opening operation: it keeps the points of S_x for all points x where the shape of S can be included within X. In other words, $\gamma_S(X)$ is given by the points of X which are covered by S when dragged inside the contours of X. In the same way as previously, these operations are *dual by complementation*:

$$X \bullet S = (X^c \circ \check{S})^c \qquad X \circ S = (X^c \bullet \check{S})^c$$

This property allows us to understand the closing from the opening. In fact, closing a set would be like adding points where $\check{S}$ cannot go, by remaining in the complement of X. It is important to note that the origin of the structuring element S no longer matters for these two operations, only the shape of the structuring element is important. Indeed, it slides inside X for opening or inside X^c for closing. While for a dilation, it is the origin of S that thins along the X border, and for an erosion, the origin of $\check{S}$ that slides along the X^c border. These operations are represented in Figure 1.4 by using the same X and S as before. Thus, for the opening represented in Figure 1.4a, the triangle which is the structuring element must remain within X, the boundary of the opening is the black curve which is contained in X. For the closing that is represented in Figure 1.4b, $\check{S}$ must remain in X^c and the boundary of the closed set is also represented in black.

These two operations have important properties because they are idempotent, that is to say:

$$\gamma_S \circ \gamma_S = \gamma_S \qquad \phi_S \circ \phi_S = \phi_S$$



(A) Opening of X by S (B) Closing of X by S

FIGURE 1.4: Example of an opening and a closing of X by S.

This property define a morphological filter, because it is necessary to apply an opening or a closing only once. It is possible to interpret the consequences of an opening: it smoothes the shapes, eliminates the connected components smaller than the structuring element, does not (always) preserve the topology and reduces the object, γ_S is *anti-extensive*. In the case of a closing, it fills the holes smaller than the structuring element, does not (always) keep the topology, welds close shapes and enlarges the object, ϕ_S is *extensive*. Moreover, the following inclusions are always true:

$$\forall X, S \in \mathcal{P}(E), \quad \gamma_S(X) \subseteq X \subseteq \phi_S(X)$$

Figure 1.5 graphically illustrates the link between the four operations we have just defined. The idempotence of closing and opening is clearly visible by following the arrows. As Figure 1.3 shows, the notion of complementation is generalized to our new operations.

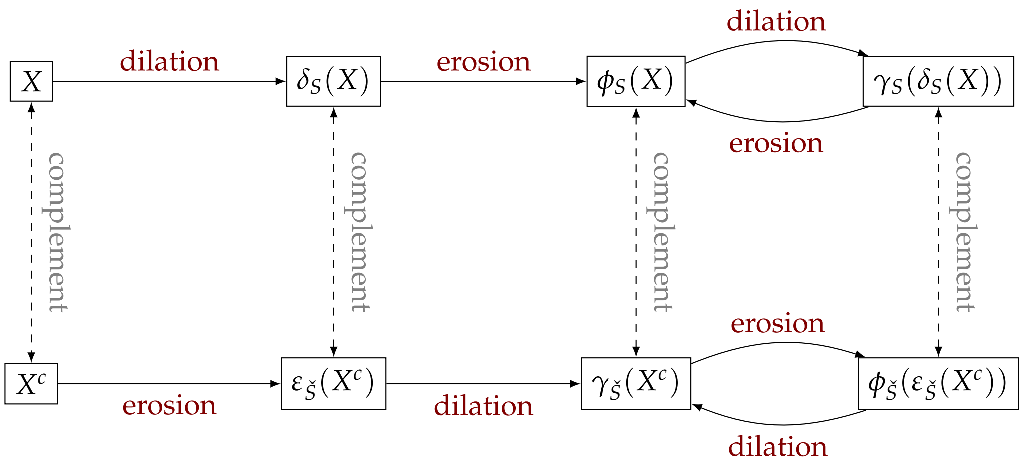


FIGURE 1.5: Graph illustrating the link between basic operations.

The four operations that were presented - dilation, erosion, opening and closing - form the basis for possible transformations in mathematical morphology. It is possible to define many other operations such as the gradient or the Laplacian by composing or adding these four basic operations.

1.1.3 Hit-or-Miss Transform

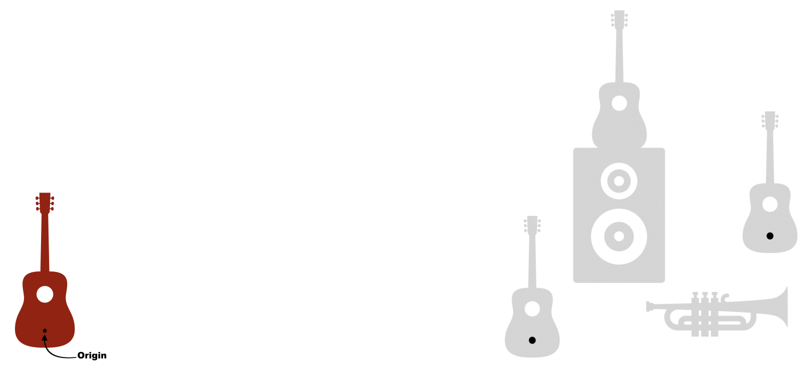
The *hit-or-miss transform*, noted by η, uses two structuring elements: S_1 the *foreground structuring element* and S_2 the *background structuring element* such as $S_1 \cap S_2 = \emptyset$. Thus, $\eta_{S_1,S_2}(X)$ contains all the points for which S_1 is contained in X and S_2 is contained in X^c. Mathematically, it is written by:

$$\eta_{S_1,S_2}(X) = X \otimes (S_1, S_2) = (X \ominus S_1) \cap (X^c \ominus S_2)$$

It is possible to define this operation with a single structuring element by setting $S_2 = S_1^c \cap X$. This is what will be used most of the time, in this case:

$$\eta_S(X) = X \otimes S = (X \ominus S) \cap (X^c \ominus S^c)$$

Since the hit-or-miss operation is formed exclusively from erosion, the origin of the structuring element must be taken into account. Indeed, the hit-or-miss operation will allow to detect patterns similar to the structuring element. These patterns will be indicated by a point which is the origin of the structuring element. Therefore, hit-or-miss transform is very useful to search for specific patterns in an image. Figure 1.6 illustrates an example of the hit-or-miss transformation. The structuring element, presented in Figure 1.6a is a classical guitar, and its origin is on the sound board. The X set in grey, in figure 1.6b, contains patterns similar to the structural element. The hit-or-miss transform will detect the patterns that look exactly like the structural element, so there are only two guitars detected. The results of the operation are in black and correspond to the origins of the two detected patterns.



(A) Structuring element S (B) Hit-or-miss (black) of X (gray) by S

FIGURE 1.6: Example of a hit-or-miss transformation of X by S.

1.1.4 Other Possible Operations

The *thickening* operation of X, $Th(X)$ is obtained by adding the result of hit-or-miss at the initial set:

$$Th(X) = X \cup \eta_{S_1,S_2}(X)$$

If the origin is contained in S_2 the background structuring element, $X \otimes (S_1, S_2) \subseteq X$ and the thickening make sense. Moreover, $th(X)$ correspond to *thinning* a set X, it is equivalent to delete the result of the hit-or-miss transform from the original set:

$$th(X) = X \cap \eta_{S_1,S_2}(X)^c = X \setminus \eta_{S_1,S_2}(X)$$

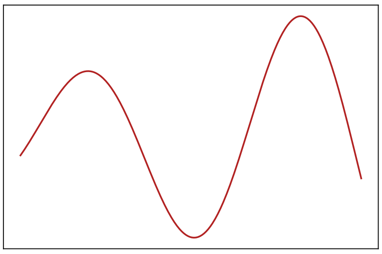
If the origin is contained in S_1 the foreground structuring element, $X \otimes (S_1, S_2) \subseteq X^c$ and the thinning make sense. It is possible to define other operations by deleting the erosion or the opening from the initial set. Thus, the *top-hat transform* τ_S is obtained by eliminating the opening of X from X:

$$\tau_S(X) = X \setminus \gamma_S(X)$$

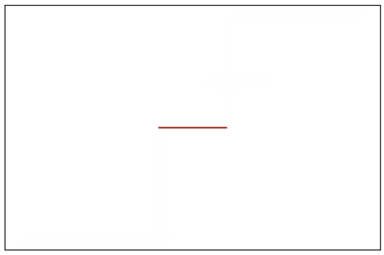
In the same way the *outline* of X is the global set X removed from the erosion of X.

1.2 Mathematical Morphology on Functions

The notions that have been mentioned until now apply to binary sets, one case of concrete applications is binary images where each pixel is black or white. To generalize these notions to grey level images, it is necessary to work with pixels that take a value that is no more binary but a number that will characterize the grey level, for example the high values will be the darkest pixels and vice-versa. Therefore, it is necessary to work with *functions* $E \longrightarrow T$ where E is the pixel space of the image and $T \subseteq \mathbb{R}$ the grayscale values. Let us take $F : E \longrightarrow T$ which will play the role of *grey-level image* and $G : E \longrightarrow T$ the *structuring function*, which respectively replace the sets X and S of the previous part. These functions having value $-\infty$ outside a bounded support. To illustrate the main possible operations, the two functions F and G will be represented in Figure 1.7. The structuring function G will be chosen constant and the origin will be placed in the middle.



(A) Grey-level image F



(B) Structuring function G

FIGURE 1.7: Example of two grey-level functions F and G.

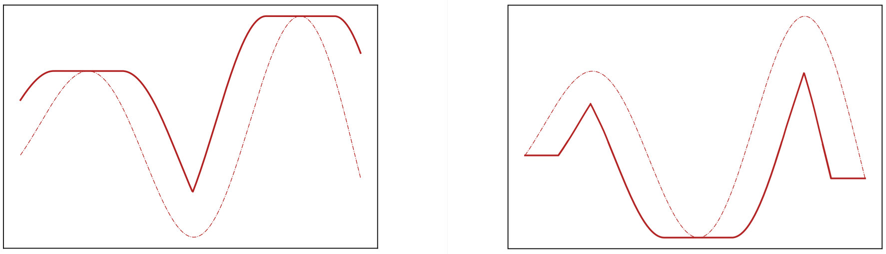
It is possible to mathematically express the *dilation* $\delta_G(F) = F \oplus G$ and *erosion* $\varepsilon_G(F) = F \ominus G$ of the F and G functions for any $x \in E$:

$$F \oplus G(x) = \sup_{t \in E} (F(x-t) + G(t)) = \sup_{t \in supp(G)} (F(x-t) + G(t))$$

$$F \ominus G(x) = \inf_{t \in E} (F(x+t) - G(t)) = \inf_{t \in supp(G)} (F(x+t) - G(t))$$

With the conventions that if in the expression $F(x-t) + G(t)$ there is ∞ or $-\infty$ the result will be equal to $-\infty$, just as with the expression $F(x+t) - G(t)$, if there is ∞ or $-\infty$ it will take the value ∞. Figure 1.8 illustrates these two operations by using the examples of the two functions in Figure 1.7. Since the origin is in the middle of the structuring function G level there is still:

$$\forall x \in E, \quad \varepsilon_G(F)(x) \leq F(x) \leq \delta_G(F)(x)$$



(A) Dilation of F by G

(B) Erosion of F by G

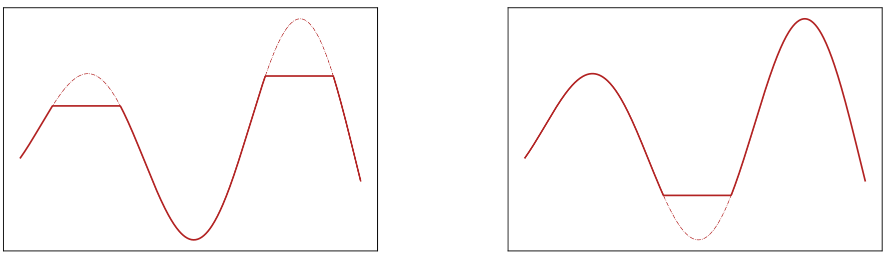
FIGURE 1.8: Example of a dilation and an erosion.

Once dilation and erosion have been defined, it is possible to define the *opening* $\gamma_G(F) = F \circ G$ and the *closing* $\phi_G(F) = F \bullet G$ in the same way as in the previous section, i.e.:

$$F \circ G = (F \ominus G) \oplus G \qquad\qquad F \bullet G = (F \oplus G) \ominus G$$

Figure 1.9 illustrates these two operations, to try to better understand these operations, let us note $Hypo(F) = \{(x, v) : x \in E, v \in T, v \geq F(x)\}$ the *hypograph* of F. It is possible to visualize the opening operation with the hypograph. In fact, the hypograph of $F \circ G$ will be the result of the opening of the hypograph of F by the graph of G, from a morphological set point of view defined in the previous section. Similarly for the $F \bullet G$ hypograph, it can be seen as the result of the closing of the F hypograph by the graph of G. As in the previous section, the following inequalities are always true:

$$\forall x \in E, \quad \gamma_G(F)(x) \leq F(x) \leq \phi_G(F)(x)$$



(A) Opening of F by G

(B) Closing of F by G

FIGURE 1.9: Example of an opening and a closing.

The properties of the four operations $\oplus$, $\ominus$, $\circ$ and $\bullet$ in the case of sets can be extended to functions. For example, the idempotence of opening and closing is preserved. These four operations on the set of function $E \longrightarrow T$ is what is called *Grey-Level Morphology* and will allow us to process and analyze images in the following sections.

Chapter 2

Mathematical Morphology on Sets Applied to Music

A video illustrating the musical examples of this chapter is available here.

2.1 Applications to Symbolic Music

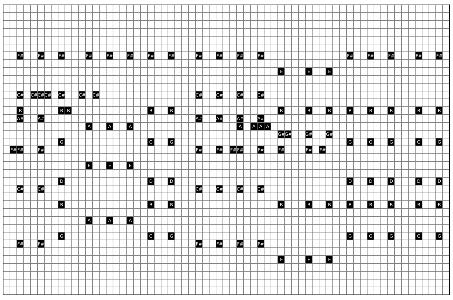
Mathematical morphology is a theory that is generally applied to image processing. Non-inversible and non-linear sets of operations modify the shapes contained in the image. They allow to filter the image by using a predefined structuring element. This theory was first developed for binary images and then extended to gray scale and color images. There are very few direct applications of mathematical morphology to music. M. Karvonen [9, 10] has used morphological tools to detect patterns in a database that are almost identical to a standard pattern. For this purpose, morphological operations are applied to a piano roll representation. Also, C. Agon et al. [11] applied morphology by working on formal concept lattices built on musical intervals.

To apply morphology to music, we will start by using morphology concepts on binary images. To do this, it is necessary to find a good representation of the music, i.e. reduce a piece of music to a binary image without losing too much information. There are many ways to do this. A first intuitive idea is to perceive the music piece as a particular piano roll. Indeed, in this piano roll, we do not take into account the duration of the notes. To process data more easily, only the apparition of a note will be important. In addition, since the image must be binary, the coefficients must be 0 or 1. The value 1 will mean that a note is played while 0 is similar to the absence of a note. Therefore, the notes are all of constant intensity. This first representation helps to understand how to apply tools from mathematical morphology to music. One of the fundamental problems is to find a good musical representation to apply the operations resulting from mathematical morphology. The piano roll presented in Figure 2.1 is a first idea, but the musical forms that appear are not compact and not connected. Since morphological operators work very well on images that contain connected shapes, efforts must be made to modify the operators and/or modify the musical representation that is currently the piano roll. Thus, the chapter 2 focuses on the classical visualization of music in the form of piano roll where it is possible but more complicated to adapt mathematical operators.

2.2 Opening Operation in Order to Extract Data

2.2.1 A First Application of an Opening Operation

In the four basic operations of mathematical morphology, the opening seems to be the most appropriate for detecting musical patterns. Indeed, this operation will allow to extract the patterns of the piano roll which are similar to the structuring element, regardless of the pitch of the note and the time at which they appear. Thus, by choosing a good structuring element, we will be able to isolate certain properties, keeping only certain chords, extracting some particular patterns or removing the melody from the piece of music. To illustrate this, let us take a fifth chord as a structuring element represented in Figure 2.1b, the note that corresponds to the fifth is seven semitones higher than the fundamental note. For this structuring element, both notes are played at the same time. Therefore, doing an opening operation by this structuring element will isolate all the fifth chords (when the fifth is played at the same time as the fundamental). The piano roll of the Pyramid Song theme of the British group Radiohead is represented in Figure 2.1a.



(A) Piano roll of Pyramid Song from Radiohead

(B) The structuring element is a fifth chord

FIGURE 2.1: Example of a piano roll and a structural element.

The result of the opening on Pyramid Song by a fifth chord is shown in Figure 2.2. The piano roll becomes much clearer because the fifth chords are clearly visible. Moreover, the melody is no longer present, we have kept only the basis of the theme. Thus, the opening operation allowed us to isolate the chords of the theme.

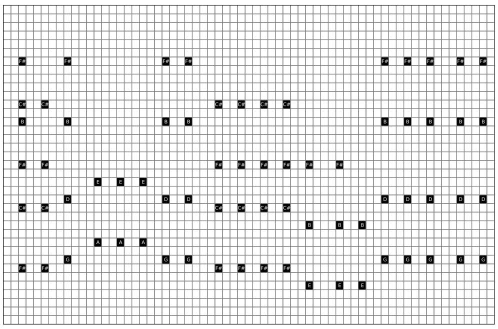


FIGURE 2.2: Piano roll of an opening by a fifth chord of Pyramid Song from Radiohead.

2.2.2 Opening with a More Relevant Structuring Element

It is possible to choose more complex structuring elements such as major, minor or seventh chords to capture further information about a piece. An opening operation of X with several structuring elements $S_1, S_2, ..., S_n$ will be defined as:

$$\gamma_{S_1,S_2,...,S_n}(X) = (X \circ S_1) \cup (X \circ S_2) \cup ... \cup (X \circ S_n)$$

The union of openings is an opening (in a general algebraic sense, i.e. increasing, idempotent and anti-extensive operator). Figure 2.3 shows the first eight bars of the famous Beatles song Hey Jude. The melody is present on the upper part of the piano roll because it is played in the high notes, while the basses are played in the lower parts. On the central part are represented the chords that are played during the piece.

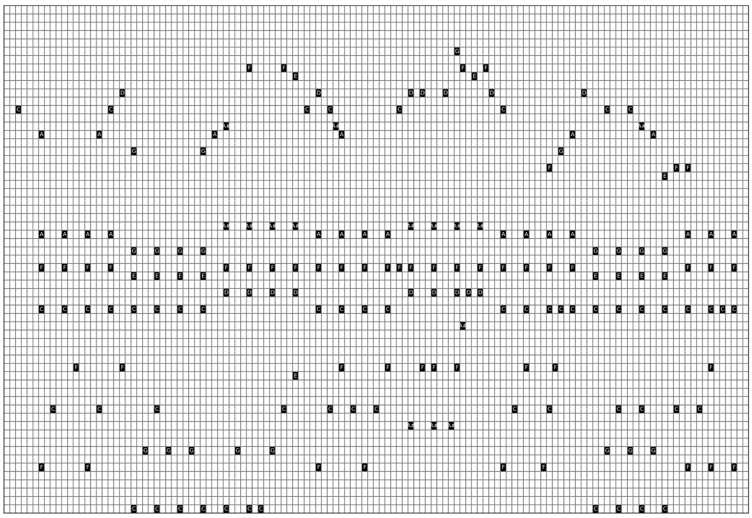
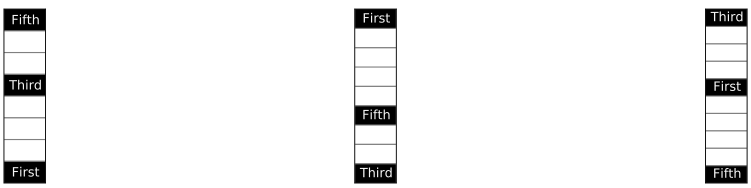


FIGURE 2.3: Piano roll of Hey Jude from the Beatles.

These measures are essentially based on major chords, so it is natural to choose a major chord as a structuring element to be able to extract the central part. Since a major chord is composed of three First/Third/Fifth notes, it can be represented in three different ways if the order is maintained, as in Figure 2.4.



(A) First/Third/Fifth (B) Third/Fifth/First (C) Fifth/First/Third

FIGURE 2.4: Structuring elements are the three representations of a major chord in the piano roll.

Taking these three representations of a major chord as structuring elements, we extract all the major chords from these first eight bars. The result of the opening is shown in Figure 2.5. Thus, only the chords are kept on this new piano roll. The piano roll is not displayed in its entirety to save space in this report but it is worth imagining that the rest of the piano roll is empty. The melodic part and the bass part have been deleted by the opening action.

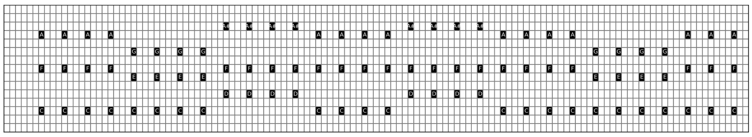


FIGURE 2.5: Piano roll of an opening by the three representations of a major chord of Hey Jude.

Once again, the structuring element is chosen so that all the notes of the chord are played at the same time. One could imagine a structuring element where the chord is spread over time like arpeggios.

2.3 Hit-or-Miss Transform to Detect the Key

2.3.1 A First Approach to Detect the Key

To analyze a piece of music, it may be necessary to know its key. Indeed, this information makes it possible to understand which scales and notes are used. For example, if the key is C major or A minor, the notes mainly used will be the seven white notes of the piano while the other five black notes of the piano will be much rarer or even non-existent in most cases. Thus, the study of a piece can generally be reduced to the study of the seven notes of the key. The Figure 2.6 highlights the notes used when the key is C major or A minor. The following pattern appears: tone/tone/semi-tone/tone/tone/tone/tone/semi-tone. This pattern is independent of the key because if the key changes it amounts to rotate the pattern in Figure 2.6. Therefore, to determine the key of a piece, it is necessary to be able to

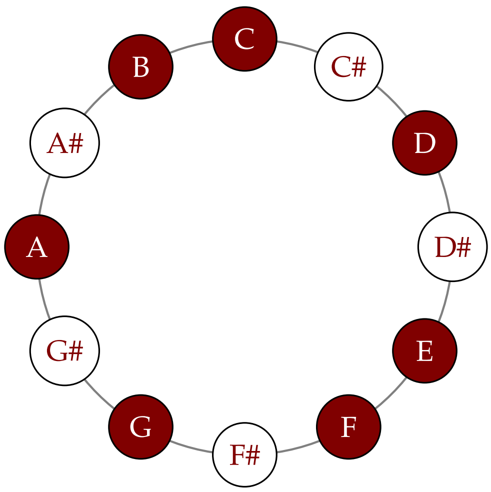


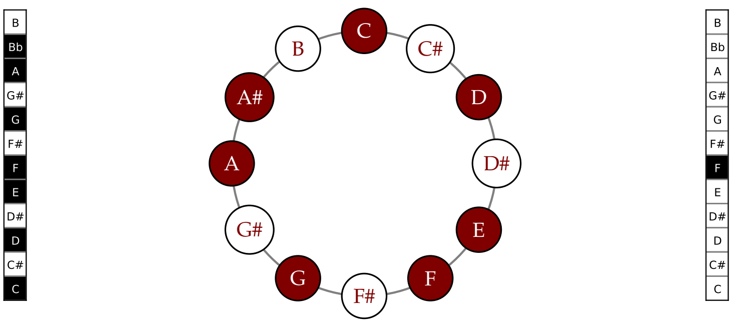
FIGURE 2.6: Twelve musical notes represented in a circle.

determine how this pattern is placed. This is exactly what the hit-or-miss operation can do. This operation makes it possible to indicate the presence of a certain pattern by its center, the center of the pattern is therefore to be chosen before performing the operation. The structuring element presented in Figure 2.7 is based on the tone/tone/semi-tone, tone/tone/tone/tone/tone/tone/semi-tone pattern. It is necessary to choose the origin of the structuring element on the fundamental, starting from the bottom, it will be the first note for the major mode in Figure 2.7a and the sixth note for the minor mode in Figure 2.7b. Thus, it is necessary to represent all

(A) Major tonality (B) Minor tonality

FIGURE 2.7: Structuring element of a major or minor key.

the notes that are played to see the pattern of Figure 2.7 appear. To do this, the piano roll must be flattened on the pitch axis and quotiented by the octave. Since this pattern contains twelve high notes, it is not enough to represent the twelve notes. The representation in the circle will allow us to have a single solution by applying the hit-or-miss transformation. Figure 2.8a represents the notes that are played during the introduction of "Hey Jude" represented in Figure 2.3. Black notes are those that are played during the MIDI file of "Hey Jude". After applying the hit-or-miss operation



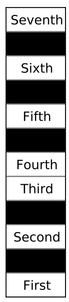
(A) Notes played (B) Notes played represented in a circle (C) Result of hit-or-miss

FIGURE 2.8: Example of a tone detection by hit-or-miss transform.

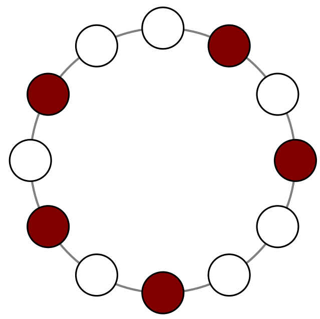
with the structuring element in Figure 2.7a in the circle in Figure 2.8b, we obtain Figure 2.8c which allows us to conclude that the key is F. It is possible then to reduce the study of the piece to the notes of the major key F which are: $C/D/E/F/G/A/Bb$.

2.3.2 A More Advanced Way to Detect the Key

However, the previous method works when all notes in the scale are played. If one of the notes is missing, which is possible if the study is performed over a short period of time, the pattern in figure 2.7a will not be detected. To remedy this, it is necessary to look at the absence of notes, that is to say to choose the complementary of the structuring element of Figure 2.7. The new structuring element is represented in figure 2.9. The origin of this structuring element will always be located in the first in order to return the key. It can be seen that this representation corresponds



(A) Structuring element



(B) Structuring element representated in a circle

FIGURE 2.9: Structuring element that will make it possible to detect the absence of notes.

exactly to the traditional image of a piano. This is quite normal because the black notes of a piano correspond to the notes that are not in the *C* major key, it is exactly the pattern formed by these notes that we will look for. Remember that this method works on MIDI files with little information, some measures with few notes. For example, the introduction to Radiohead's Videotape that is played on the piano is represented as a piano roll in Figure 2.10. The flattened piano roll is shown in Figure 2.11a. There are few different notes played: $C\#/A/G\#/F\#/E$, not enough to fill all the notes of a key. Therefore, the previous method will not work. It is necessary to

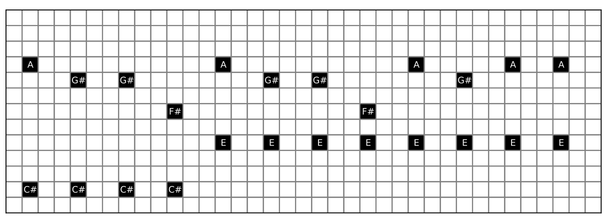
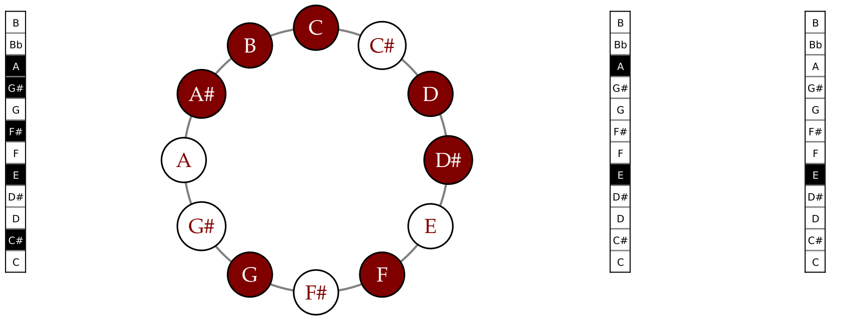


FIGURE 2.10: Piano roll of the theme of Videotape from Radiohead.

be interested in the notes that are not played, for that it is necessary to work with the complementary of the flattened piano roll which appears in Figure 2.11b represented in a circle. Among the notes in the circle that are not played, look for the pattern of Figure 2.9b. To do this, it is also necessary to apply the hir-or-miss operation with this structuring element. The pattern appears several times when there are not many notes in the MIDI file as in the example chosen. In our case, the pattern appears twice in Figure 2.11c. This means that there are two possible choices for the tone. It is then necessary to choose among these different possible tones. In our example in Figure



(A) Piano roll (B) Complementary in a circle (C) Hit-or-miss (D) Tonality

FIGURE 2.11: Example of a tone detection by the hit-or-miss operation by using the complementary method.

2.11c the two possible tones are *A* and *E*. However, it is easy to notice on the piano roll in Figure 2.10 that the notes *C#* and *E* are predominant. This argument leads to the conclusion that the key will be *E* and not *A*. In conclusion, by observing the frequency at which notes appear during the MIDI file, it allows one to make a choice to determine the right key.

2.3.3 Representation in Relation to the Key

Once the key of a piece has been determined, it is possible to represent the piano roll in this key. In the previous example in Figure 2.10, we determined that the key is *E*. Figure 2.12 represents the piano roll in the key of the song. In general, a tonality representation is equivalent to deleting five notes on the piano roll, which means going from twelve to seven notes. The result is a more compact representation. Another important property that results from a representation in key is that major and minor chords are represented by the same patterns. Indeed, in the tonality, there will always be only one note between the first and the third and one note between the third and the fifth. Consequently, major or minor chords will be in a single form, with the permutation of the three notes keeping the order, so three possible patterns, which are illustrated in Figure 2.13.

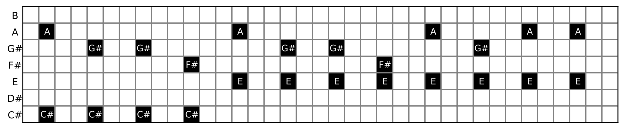
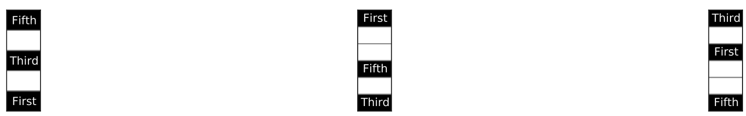


FIGURE 2.12: Piano roll immersed in its tonality of the theme of Videotape from Radiohead.



(A) First/Third/Fifth (B) Third/Fifth/First (C) Fifth/First/Third

FIGURE 2.13: Representations of a major/minor chords in the piano roll.

Chapter 3

A New Approach for Visualizing Musical Pitch Variations

A video illustrating the musical examples of this chapter is available here.

3.1 Identify Melodic Patterns in a Melody

3.1.1 Description of the Method for a Melody

The objective of this chapter will be to characterize the melodies by detecting certain repetitive patterns. These patterns are the result of pitch changes within the melody. In a completely natural way, let us represent the notes of the melody on a graph where the abscissa axis represents time while the ordinate axis represents the pitch of the note. Thus, a note played at time t at a pitch p will be the point (t, p). This note will be linked to the next one to form the *Melodic Curve* which characterize the melody. The *Melodic Contour* is then defined by the consecutive normalized slopes of this curve. That is too say, the melodic contour is defined by the set of direction between consecutive pitches of a melody, +1 and -1 indicating respectively an ascending and a descending interval, while 0 indicates that the interval is neither ascending nor descending. Figure 3.1 gives an example of a melodic contour equal to $\{+1, 0, -1, +1, -1\}$.

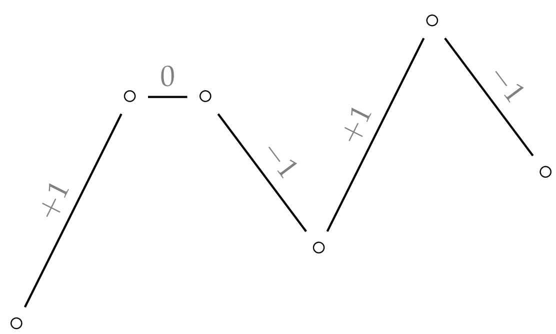


FIGURE 3.1: The melodic contour is equal to {+1, 0, -1, +1, -1}.

The melodic contour has become a fundamental tool whether for music perception [12, 13], music analysis [14] or music theory [15]. In our case, it will be used to simplify the information and to characterise musical patterns. By studying the melody over a much longer period of time, it is possible to see some patterns appear in the melodic contour. The next few pages of this report will focus on this type of pattern, called *Melodic Pattern*, in two famous musical works and then generalize this notion.

3.1.2 Prelude of Bach in C major, BWV 846

One of the most famous pieces of classical music is Johann Sebastian Bach's first prelude published in 1722 . This is the major *C* prelude that was later classified as BWV 846. This piece is the introduction to the book The Well-Tempered Clavier, which contains a prelude and a fugue in each of the twelve semitones in major and minor, i.e 48 works. The melodic curve at the beginning of the piece is shown in Figure 3.2.

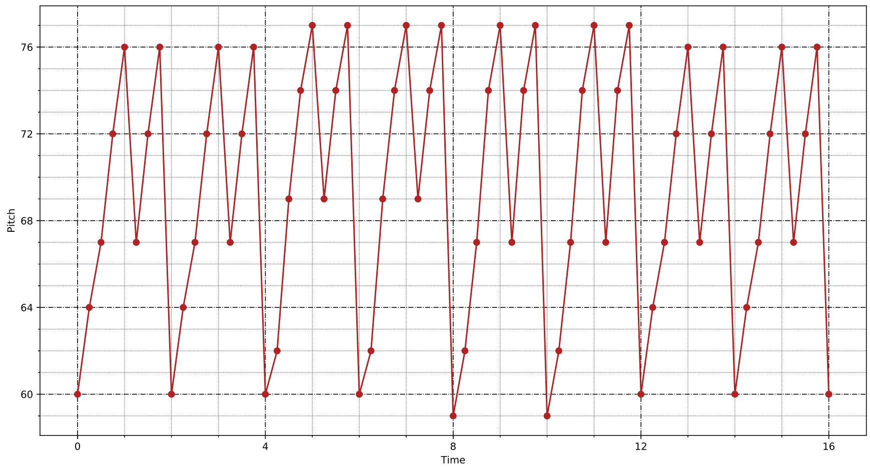


FIGURE 3.2: Melodic curve of BWV 846.

To simplify the visualization, the melodic contour is represented as an image in Figure 3.3, a black pixel will mean a value of $+1$, a white pixel -1 and grey pixel 0. This representation is much more explicit and makes it easier to identify melodic patterns. As with the melodic curve, the melodic contour is not represented in its entirety due to lack of space. However, it is necessary to imagine that this one extends in a similar way to what is represented.

FIGURE 3.3: Representation of the melodic contour of BWV 846.

Thus, a particular pattern emerges, it will be present throughout the whole piece. This melodic pattern is represented in Figure 3.4 and its values are respectively:

$$\{+1, +1, +1, +1, -1, +1, +1, -1\}$$

This analysis works with this piece because it has a strong melodic regularity. Not all Bach's pieces are as regular, but in many situations it is possible to find a melodic pattern which is present many times and that characterizes the melodic contour.

FIGURE 3.4: Melodic pattern of BWV 846 that appears in the melodic contour.

3.1.3 Études of Chopin Op. 10 No. 1

During the 1830s, Chopin published two books of Études of twelve études, Op. 10 and Op. 25, the first of which was dedicated « à son ami F. Liszt » ("to my friend F. Liszt"). The first étude of Opus 10 contains a melodic pattern present throughout the study. This pattern played exclusively by the right hand on the piano makes this étude the most difficult to play of the two opuses. The profile of the melodic curve is represented in Figure 3.5. As before, the first eight measures are represented in this figure, but the pattern remains almost unchanged throughout this study.

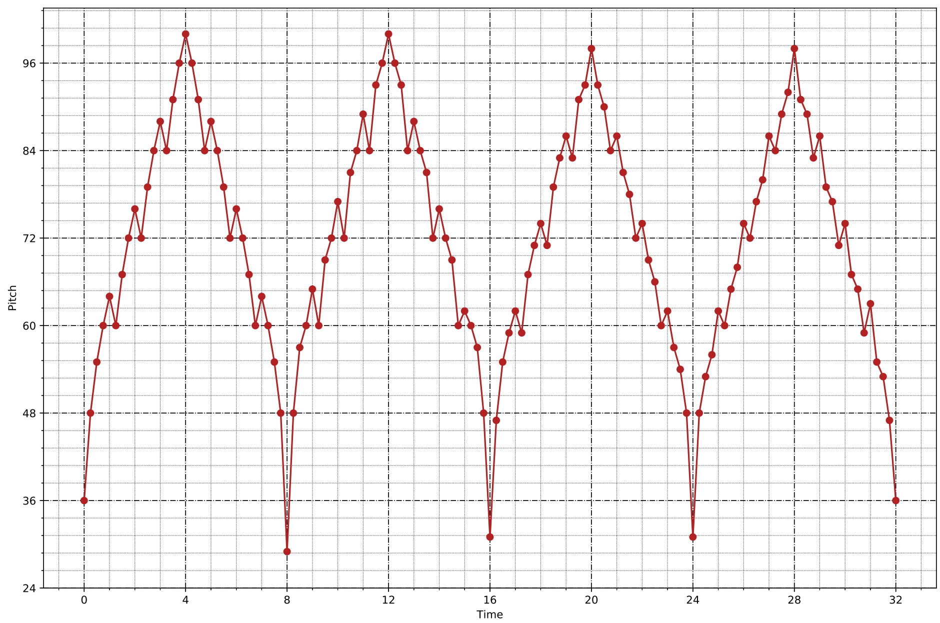


FIGURE 3.5: Melodic curve of Chopin's Étude Op. 10, No. 1.

As in the previous section, the melodic contour can be determined from the melodic curve. This one, represented in Figure 3.6, is more complex than the previous one. This is because the melodic pattern itself is longer and richer than the previous one.

FIGURE 3.6: Representation of the melodic contour of Chopin's Étude Op. 10, No. 1.

A melodic pattern emerges in the melodic contour. Represented in Figure 3.7, it contains 32 values compared to 8 for the melodic pattern of the prelude BWV 846. Moreover, this pattern has an inverse symmetry with respect to its centre, i.e. the second half of the pattern is deduced from the first half by an inverse symmetry.

FIGURE 3.7: Pattern that appears in the melodic contour.

The melodic patterns in Figure 3.4 and Figure 3.7 allow for a better understanding of the regularity of the piece compared to the traditional vision of the pattern on the musical score. Indeed, Figure 3.8 is a score that illustrates the pattern present on these two pieces. This musical score allows to compare the two melodic patterns, it is clear that they are not the same but it is complicated to extract exactly their characteristics. For this reason, the presentation in black and white squares form is much more visual and intuitive, it will allow more complex cases to be handled.

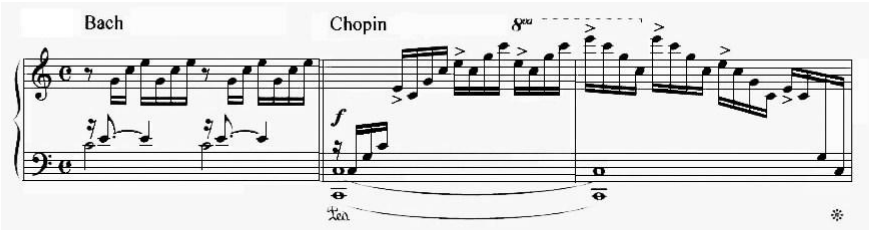


FIGURE 3.8: Comparison of Bach's Prelude No. 1 in C major with Chopin's Étude Op. 10, No. 1.

3.2 A Matrix Representation to Generalize Melodic Patterns for a Chord Sequence

3.2.1 Definition of the Transition Matrix between Two Chords

The previous section has made it possible to characterize certain regularities in the melodies by studying the variations of the pitch. Numbers $+1, -1$ or 0 were used to model these changes. Considering the importance of the melodic contour, it is not surprising that multiple extensions have been proposed. For example, two other contours were defined in [16]: the *strong contour* (melodic contour of only the notes present on the beat) and the *weak contour* (strong contour with extra information if there is a contour variation within the beat). Moreover, it was proposed in [17, 18] to observe the directions at longer range, i.e., all the directions between the i^{th} and j^{th} pitches, not only between the i^{th} and $(i+1)^{th}$ pitches as for the usual melodic contour. However, these generalizations remain in the monophonic context, and they do not handle musical chords. We propose a generalization of the melodic contour to chord sequences, i.e., not restricted to note sequences.
During a sequence of two chords, it is impossible to use the previous method because it makes no sense to say that the chord is higher or lower if several notes are played at the same time. Therefore, it is no longer a number that will characterize this sequence but a matrix that we will call *Transition Matrix*. The coefficient of the i^{th} line and the j^{th} column is the normalized value of the derivative of the curve that passes through the i^{th} lower note of the first chord and through the j^{th} lower note of the second chord. Thus if the first chord has n notes and the second m notes the dimension of the transition matrix will be (n, m). Figure 3.9 illustrates how the transition matrix is constructed, in this example, the two chords have two notes, so it will be a matrix $(2, 2)$. To know the first line, the lowest note of the first chord must be compared to the other notes of the next chord. The second line is constructed in the same way with the other note of the first chord. This new method will have a direct impact on the melodic contour. Instead of a sequence of numbers, it is now defined more generally as a sequence of matrices. Thus, the transition matrices will

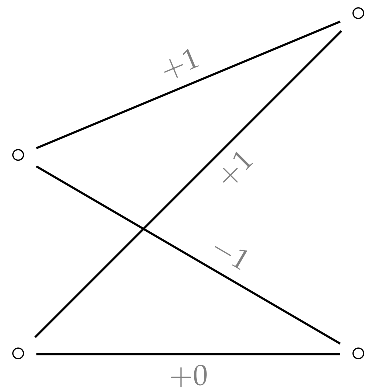


FIGURE 3.9: In this example the transition matrix is $\begin{pmatrix} +0 & +1 \\ -1 & +1 \end{pmatrix}$.

be assembled end to end to form the chord contour. Subsequently, the melodic patterns will be defined in the same way by observing the recurring patterns in the chord contour. The next two sub-sections will be devoted to the analysis of melodic patterns when the dimension of the transition matrices are constantly $(2, 2)$.

3.2.2 Prelude of Bach in C minor, BWV 847

After studying the first prelude of The Well-Tempered Clavier we can move on to the second prelude, the one in C minor noted BWV 847. This prelude is also very regular because it is all played in sixteenth notes. The fundamental difference with the first prelude is that it is a étude with two voices where each note played is doubled. In a certain sense this prelude is a two notes chord sequence. To visualize this, it is

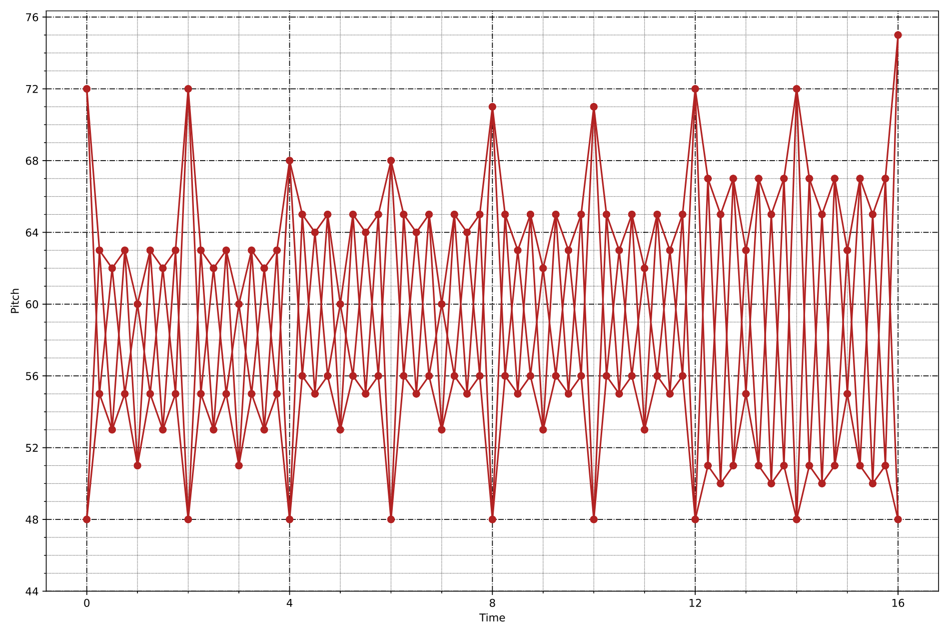


FIGURE 3.10: Melodic curves of BWV 847.

possible to link all the notes of one chord to all the notes of the next chord, which is represented in Figure 3.10. Since all transition matrices are of size (2.2), the chord contour, which represent the sequence of transition matrices, will have two lines. Figure 3.11 represents a part of the chord contour. The chord contour appear as a

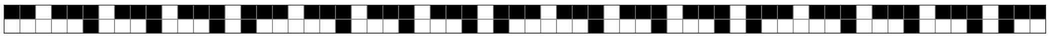


FIGURE 3.11: Representation of the chord contour of BWV 847.

"frieze" that contains a certain periodicity. The periodicity that appears in the chord contour is highlighted by the melodic pattern in Figure 3.12. This pattern is a union of eight matrices (2,2). The melodic pattern thus obtained contains two dimensions, it is a generalization of the one-voice melodic patterns studied in the previous pages.

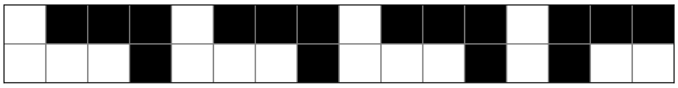


FIGURE 3.12: Melodic pattern of BWV 847 that appears in the chord contour.

Despite the lack of direct interpretations of the melodic pattern, it remains much more visual than on a musical score. Indeed, Figure 3.13 illustrates the first four measures of BWV 847. The melodic pattern is present height times on these two bars. However, it is much more complicated to see it in relation to chord contour. Thus, the interest of going through this matrix representation and the chord contour makes it much easier to detect patterns within a musical piece.

FIGURE 3.13: Musical score of the first four measures of BWV 847.

3.2.3 Godowsky's Arrangement Op. 10 No. 1

To continue the analogy with the first part, let us study again Chopin's study Op.10 No.1. More precisely, Leopold Godowsky's arrangement of Op. 10 No. 1. Studies on Chopin's Études were composed between 1893 and 1914 by Leopold Godowsky. This one added complexity to the pieces already written by Chopin. In the case of Op. 10 No. 1, the initial melody is now played on the left hand while each note is doubled by the right hand. Again, this piece is a succession of two-note chords. From a graphical point of view, the melodic curves are represented in Figure 3.14.

As before, the chord contour will have two lines because the transitions matrices are all of size (2,2). This one is much more complicated than the previous one because the melody is much more complex. The chord contour is represented in Figure 3.15,

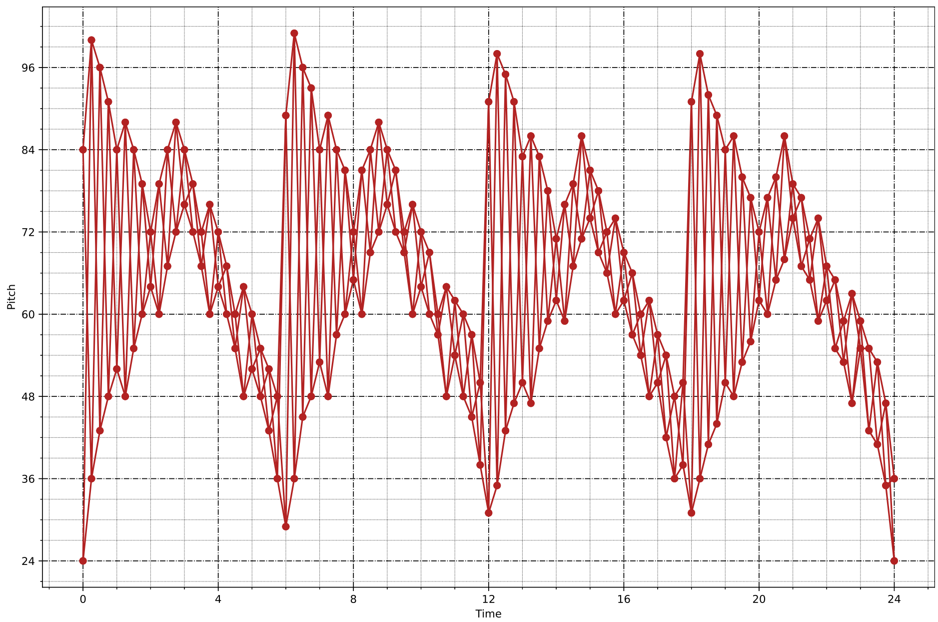


FIGURE 3.14: Melodic curves of Godowsky's arrangement Op. 10 No. 1.

this time it is represented in its entirety in relation to the melodic curves of Figure 3.14. The chord contour is not perfectly periodic, but there is a pattern that seems to appear four times. In the three previous examples, the chord contour was periodic but in many situations, like this one, the chord contour have a certain regularity that is not perfect. The chord contour will be described as *almost periodic*, because this pattern is not exactly identical every time. The almost periodicity of the chord contour will be predominant in corpus analysis in general. As for the almost periodic functions, it is necessary to set a threshold value and a distance between two melodic patterns. If the distance is less than the threshold, the pattern is considered to be fairly close to the previous one and if the chord contour respects this characteristic in its entirety it will then be almost periodic.

FIGURE 3.15: Representation of the chord contour of Godowsky's arrangement Op. 10 No. 1.

Since the piece is considered almost periodic, it is necessary to choose among several melodic patterns that are similar. Nevertheless, to better represent the piece, it is more appropriate to choose the pattern that is most often present. In the case of Godowsky's arrangement Op. 10 No. 1, the melodic pattern, composed of 24 transition matrices, which occurs most often is represented in Figure 3.16.

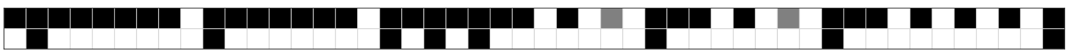


FIGURE 3.16: Melodic pattern of Godowsky's arrangement Op. 10 No. 1 that appears in the chord contour.

3.3 Theoretical Analysis of Transition Matrices

The purpose of this last section is to focus on the characteristics of the transitions matrices. First of all, if the first chord has n notes and the second m notes they will be of size (n, m) and they are of the following form:

$$\begin{pmatrix} & +1 & +1 & +1 & +1 & +1 \\ & & +1 & +1 & +1 & +1 \\ -1 & & & +1 & +1 & +1 \\ -1 & -1 & & & +1 & +1 \\ -1 & -1 & -1 & & & +1 \\ -1 & -1 & -1 & -1 & & \\ -1 & -1 & -1 & -1 & -1 & \end{pmatrix}$$

That is to say, a transition matrix is composed of +1 in the upper right corner and -1 in the lower left corner. All the information in a transition matrix is located at the diagonal level, i.e. at the intersection between these two values. With regard to the i^{th} line of the matrix, the change between the values -1 and +1 can be done in two different ways, by a 0, which means that the i^{th} note of the first chord is also in the second chord, or without intermediate value which means that the i^{th} note of the first chord is not in the second chord. The transition matrix that sends an chord of n notes on itself, called E_n, is the *neutral element* of size n and is of the form:

$$E_n = \begin{pmatrix} 0 & +1 & \cdots & +1 \\ -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & +1 \\ -1 & \cdots & -1 & 0 \end{pmatrix}$$

A surprising property of transition matrices comes from the fact that we can define an *inverse* M^{-1} of a transition matrix. Let A_1 and A_2 be two chords and M the transition matrix which goes from A_1 to A_2, the transition matrix which goes from the A_2 to A_1 is $-M^T$, therefore:

$$M^{-1} = -M^T$$

It is possible to use transitions matrices to determine whether or not one chord is higher-pitched sound than another. Indeed, in Figure 3.17, it is complicated to judge whether the chord is higher-pitched or lower-pitched than the previous one. To do this, when both chords have the same number of notes it is possible to use the *trace* of the transition matrix. When the chords have a different number of notes (or not), it is possible to look at the sum of all the coefficients of the transition matrix. In both cases, by noting Tr the result of the trace or the sum of coefficients of the transition matrix, it is possible to conclude with the following reasoning:

$$Tr > 0 \Rightarrow \text{Higher chord} \qquad Tr < 0 \Rightarrow \text{Lower chord} \qquad Tr = 0 \Rightarrow \text{To Discuss}$$

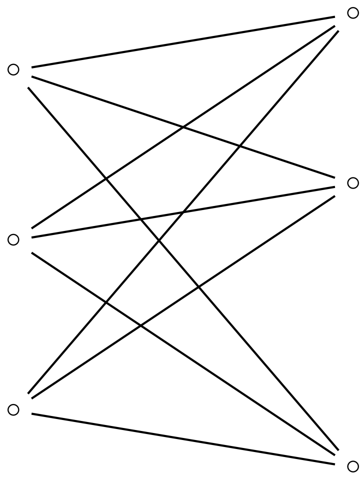


FIGURE 3.17: In this case the transition matrix is $\begin{pmatrix} -1 & +1 & +1 \\ -1 & +1 & +1 \\ -1 & -1 & +1 \end{pmatrix}$.

The i^{th} coefficient of the diagonal of the transition matrix corresponds to the slope of the melodic curve, always normalized, passing through the i^{th} note of the first chord and the i^{th} note of the second chord, therefore the diagonal identifies a direct matching between the notes of the first and second chord. Thus, in Figure 3.17, the two chords each have three notes. It is therefore possible to determine the trace of the transition matrix. The value of the trace is +1, so it is positive and this allows us to conclude that the second chord is higher-pitched than the first one. By calculating the sum of the coefficients of the matrix, which corresponds to a comparison of each note of the first chord with all those of the second chord, we also obtain the value +1 which concludes with the same result.

Chapter 4

Analysis of the Musical Structure based on Pitch Variations

Video illustrating this chapter is available by clicking here.

4.1 Introduction to Music Structure Analysis with the Study of the Chord Contour

In the previous chapter, very regular pieces were analyzed where there was only one melodic pattern. However, in the majority of music pieces there are several melodic patterns present, for example one pattern may correspond to the verse and another to the chorus etc. In addition, the size of the transition matrices varies because the number of notes varies throughout the song, so the melodic patterns will be richer and more varied. The chord contour, which is the set of transition matrices placed end to end, can still be represented graphically. Indeed, Figure 4.1 represents the introduction, approximately the first 40 seconds, of the piece at the piano *March of the Dwarfs* composed by Edvard Grieg in 1891. As before, a black pixel corresponds to a matrix coefficient of a value of $+1$, a light grey pixel to -1 and a dark grey pixel to 0. The white value that is present in the chord contour corresponds to the absence of a matrix coefficient because the size of the transition matrices varies.

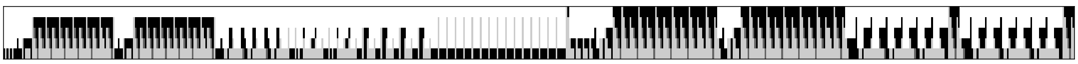
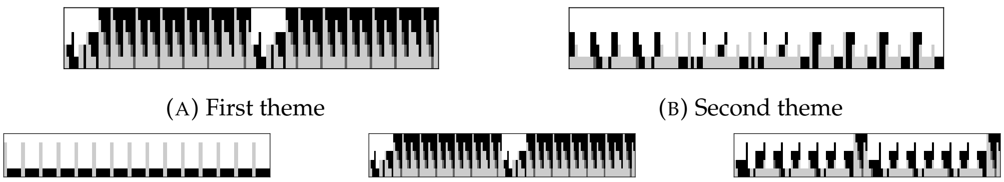


FIGURE 4.1: Chord contour of the introduction of March of the Dwarfs.

The chord contour can be divided into five passages that can be visually identified. These five passages are represented in Figure 4.2. Thanks to this analysis it is possible to start to identify the structure of this piece. In addition, the first (Figure 4.2a) and fourth (Figure 4.2d) themes are very similar.



(A) First theme

(B) Second theme

(C) Third theme

(D) Fourth theme

(E) Fifth theme

FIGURE 4.2: Chord contour of the different musical themes in the introduction of March of the Dwarfs.

The rest of this section will focus on the study of the chord contour for a piece in general. That is to say how to extract the musical structure of the chord contour and how to define a distance between the obtained patterns?

4.2 Similarity Inside the Chord Contour

To calculate the similarities within the chord contour a common method in musical analysis is to study the *Self-Similarity Matrix* [19]. To do this, it is first necessary to define a distance on the elements contained in the chord contour, that is to say on the transition matrices.

4.2.1 Distance Between Two Transition Matrices

The main difficulty in defining a distance on the transition matrices comes from the fact that these matrices can be of different sizes. Let us first consider the case where two transition matrices have the same size, the *Hamming distance* will be use. Let A and B be two transition matrices of sizes (n, m), and N the function which is null in 0 and which is equal to 1 elsewhere, i.e. $N(x) = 1$ for $x \neq 0$ and $N(x) = 0$ for $x = 0$. The distance $d(A, B)$ between the matrices A and B is defined as the number of different coefficients between these two matrices:

$$d(A, B) = \sum_{i=1}^{n} \sum_{j=1}^{m} N(a_{i,j} - b_{i,j}),$$

where $(a_{i,j})_{i,j}$ and $(b_{i,j})_{i,j}$ are the coefficients of the A and B matrices.
If one of the two matrices has more rows (or columns) than the other matrix, the idea is to delete rows (or columns) to reduce it to get two matrices of the same size and use the previous formula to which the number of rows (or columns) deleted will be added. The rows (or columns) to be deleted are those that minimize the distance between the two matrices. A row (or column) deleted from a transition matrix corresponds to a note that is deleted in the first chord (or second chord). Thus, for A and B two matrices of size (n_1, m_1) and (n_2, m_2), the distance $\mathcal{D}(A, B)$ between these two matrices of different sizes is:

$$\mathcal{D}(A, B) = \min_{\Phi,\Psi} \left(\sum_{i=1}^{\min(n_1,n_2)} \sum_{j=1}^{\min(m_1,m_2)} N(a_{\Phi(i,j)} - b_{\Psi(i,j)}) \right) + |n_1 - n_2| + |m_1 - m_2|,$$

where $\begin{cases} \Phi : \{1, ..., \min(n_1, n_2)\} \times \{1, ..., \min(m_1, m_2)\} \longrightarrow \{1, ..., n_1\} \times \{1, ..., m_1\} \\ \Psi : \{1, ..., \min(n_1, n_2)\} \times \{1, ..., \min(m_1, m_2)\} \longrightarrow \{1, ..., n_2\} \times \{1, ..., m_2\} \end{cases},$

are two strictly increasing functions, $(a_{i,j})_{i,j}$ and $(b_{i,j})_{i,j}$ are always the coefficients of the matrices A and B.

From a mathematical point of view, the first metric d respects symmetry, the identity of indiscernibles, the non-negativity and the triangular inequality so it is well defined as a metric on the matrix space with the same size. On the other hand, for the second distance $\mathcal{D}$, symmetry, the identity of indiscernibles and the non-negativity are respected but not the triangular inequality, which does not really define a metric in the mathematical sense but a semimetric. However, since it is only necessary to make comparisons between two matrices, without looking for a path from one matrix to another, triangular inequality is not essential.

4.2.2 Self-Similarity Matrices Based on the Chord Contour

Using the semimetric $\mathcal{D}$ defined previously it is possible to determine the Self-Similarity Matrix of the chord contour. By noting n the number of transition matrices contained in the chord contour, the size of the self-similarity matrix is (n, n). By noting M_k the k^{th} transition matrix of the chord contour, the coefficient $C_{i,j}$ of the line i and the column j of the self-similarity matrix is defined by:

$$C_{i,j} = \mathcal{D}(M_{n-i}, M_j)$$

Since $\mathcal{D}$ is symmetric the self-similarity matrix will be symmetric with respect to the antidiagonal (which is the diagonal that starts from the bottom left corner and goes to the top right corner). Moreover, the antidiagonal is very useful because by identifying the different blocks, it allows us to detect the zones which are similar to themselves so the different musical passages. Therefore, the antidiagonal allows us to have a direct information on the structure of the piece and to identify the different melodics patterns. Figure 4.3 represents the self-similarity matrix using the example of the introduction of March of the Dwarfs where the chord contour are represented in Figure 4.1. The value equal to zero is symbolized by the white color while the black color means a high value of the semimetric $\mathcal{D}$.

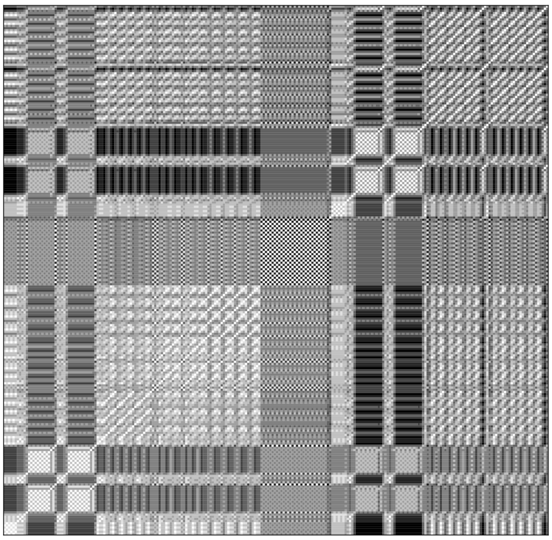


FIGURE 4.3: Self-similarity matrix of the introduction of March of the Dwarfs.

The different passages can be read on the antidiagonal and five passages can be visually identified. The similarity between the first and fourth themes is clearly perceptible on the antidiagonal of this matrix. These passages have been highlighted in Figure 4.4, it is possible to detect them manually. In the following section, we will use methods to filter this matrix to be able to detect these passages automatically.

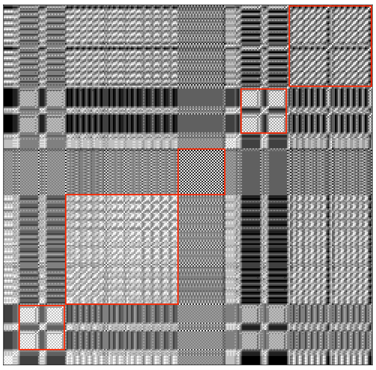
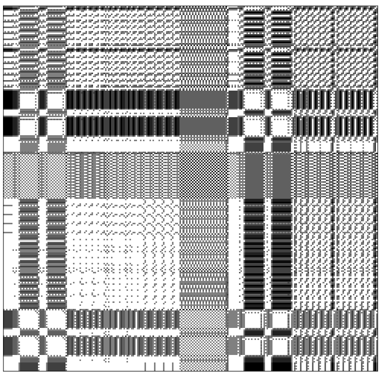


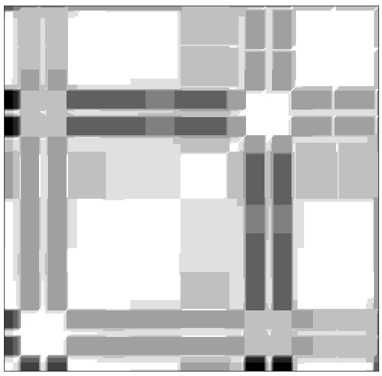
FIGURE 4.4: Different themes on the antidiagonal of the self-similarity matrix.

4.3 Filtering of the Self-Similarity Matrix using Morphological Tools to Determine the Structure of the Piece

To extract data from the self-similarity matrix, it is necessary to isolate the blocks on the antidiagonal, i.e. find the structures identified in Figure 4.4. The objective will be to homogenize the areas of the self-similarity matrix. For this purpose, it is possible to use morphological tools. Opening is particularly well adapted to this situation. Since all the antidiagonal coefficients are zero, because: $\forall i \in \{1, ...n\}, C_{n-i,i} = \mathcal{D}(M_i, M_i) = 0$, the blocks of the self-similarity matrix represented in Figure 4.4, that are on the antidiagonal will be reduced to the smallest local value, that is 0. By taking a constant and square structuring element, because we want to keep the angles of the blocks, it is then possible to homogenize the areas of the self-similarity matrix, and even stronger to reduce the blocks that are on the antidiagonal to a zero value. Figure 4.5 illustrates two filtration techniques, the first method, in Figure



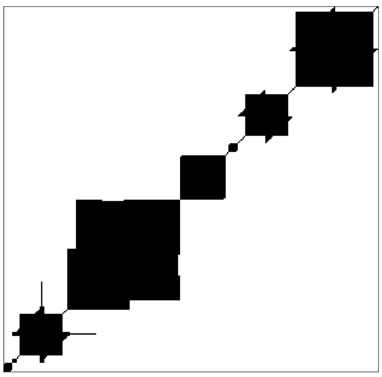
(A) Threshold applied on the self-similarity matrix



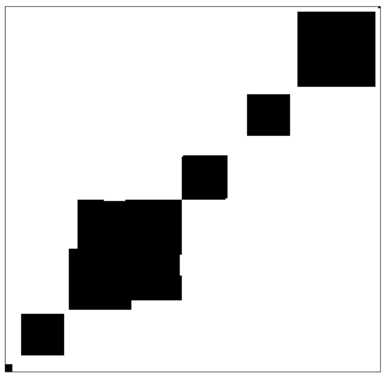
(B) Opening applied on the self-similarity matrix

FIGURE 4.5: Morphological filters applied to the self-similarity matrix to extract information on the song structure.

4.5a, is frequently used, it is a threshold filtering. This technique is very simple and consists in defining a threshold value below which each coordinate of the matrix is reduced to zero. The second method, illustrated in Figure 4.5b, is the method we presented: opening by a constant and square structuring element. The goal is to detect the overall structure of the piece, so the large main blocks of the antidiagonal, so it is necessary to choose a sufficiently large structuring element, that is 12x12 in size. The thresholding does not allow us to homogenize the blocks while the antidiagonal blocks are well reduced to zero (in white color) with the opening. Indeed, thresholding is an operation that acts globally on the matrix, the same threshold value is applied everywhere. On contrary, opening is an operator that acts locally on the zones of the matrix, as if the threshold value were to change depending on where the operation applies.
The flood fill algorithm allows us to isolate a connected component. Thanks to this, it is possible to isolate the antidiagonal from the Figure 4.5b. The connected component contains the different blocks that we are looking for because with opening they have all been set to a zero value, so they form a single related component. Figure 4.6a represents the connected component of the antidiagonal. This figure can be made clearer by applying an opening to this new image again, which is now a binary image and therefore no more a grayscale image. The final result is shown in Figure 4.6b. In this last figure, it is clearly possible to identify the five blocks searched.



(A) Flood fill applied to the antidiagonal of the self-similarity matrix



(B) Opening applied on the connected component to isolate the five blocks

FIGURE 4.6: Analysis of the information contained around the antidiagonal of the self-similarity matrix.

The structure of the piece is generally identified. Nevertheless, it is legitimate to ask whether it is possible to detect the song structure on a smaller scale to identify the bars, and therefore to detect shorter melodic patterns [20]. This can be done by changing the size of the opening structuring element applied to the self-similarity matrix. Indeed, by taking a structuring element that is always homogeneous and square but this time smaller, it is possible to detect blocks around the antidiagonal that will be much smaller. To summarize, by changing the size of the structuring element, we modify the precision of the block detection. A small structuring element will detect melodic patterns contained in a bar while a large structuring element will detect melodic patterns of a longer duration. This idea will be developed in the next section.

4.4 Detection of Blocks Around the Antidiagonal of the Self-Similarity Matrix at Several Levels of Filtering

To illustrate the notion of filtering with increasingly structuring elements to detect larger melodic patterns, we will study the famous *Alla Turca* composed in 1780 by Wolfgang Amadeus Mozart and popularly known as the *Turkish Rondo* or *Turkish March*. The objective is always to find melodic patterns that allow us to characterize musical themes. These melodic patterns are still to be found inside the chord contour. Given the length of the piece, it is preferable to represent the chord contour in different parts that is why it is cut in three in Figure 4.7. Directly identifying the different melodic patterns on this figure is not obvious because there is much more data, so finding similarity to the eye becomes more complicated. The structure of the

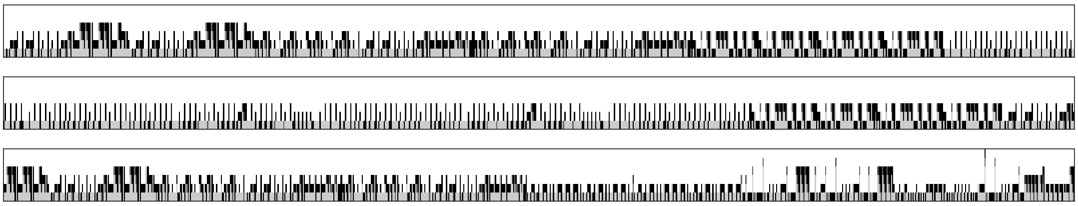


FIGURE 4.7: Chord contour of Alla Turca.

piece is shown in Figure 4.8. This piece is divided into four main parts represented by grey rectangles. These four parts are linked by the theme numbered by the letter C and C'. The first and third parts are almost identical and contain the musical themes A and B. The second part has a similar structure to the first and third part but with the musical themes D and E, while the last final part is symbolized by the letter F. Since the chord contour is much longer than before because the length of

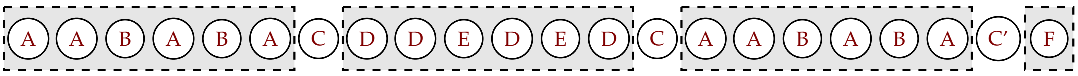


FIGURE 4.8: Musical structure of Alla Turca.

the piece is, the self-similarity matrix is also larger because it has more data, this one is represented in Figure 4.9. Let us start by trying to detect the global structure of

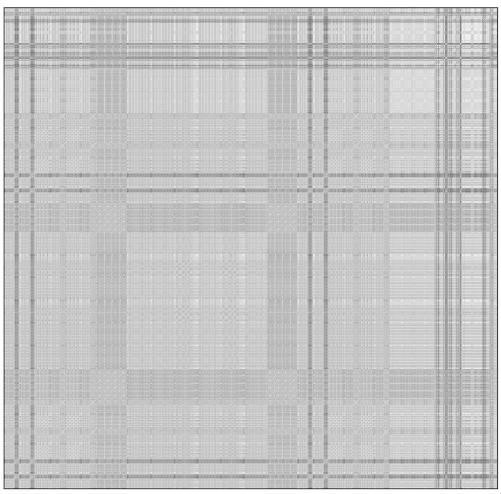
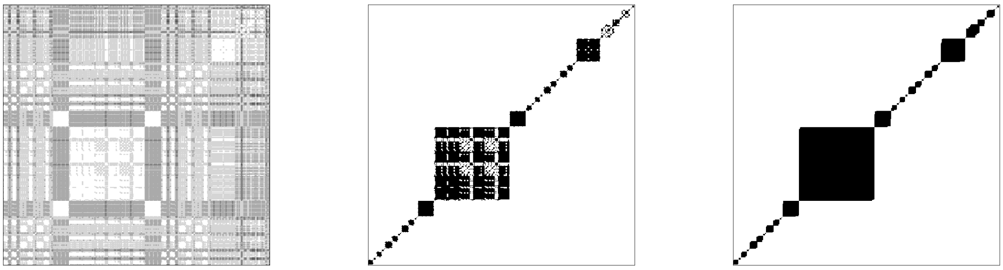


FIGURE 4.9: Self-similarity matrix of Alla Turca.

the piece, i.e. detect the four parts that are represented by grey rectangles in Figure 4.8 and the three transitions that are represented by the letters C and C'. To do this, the self-similarity matrix must be filtered through a large structuring element to lose as much detail as possible by homogenizing the blocks. Let us take a structuring element which is a square of length 6 where all the coefficients are equal to 1, this element is considered to be of high size compared to the size of the self-similarity matrix. The result of the opening on the self-similarity matrix in Figure 4.9 by this structuring element is presented in Figure 4.10a. The matrix seems more exploitable, but it is by isolating the antidiagonal using the flood fill algorithm that we begin to understand the structure of the part, this is represented in Figure 4.10b. Nevertheless, there is a frequent problem, it is necessary to apply an additional processing to the antidiagonal. Indeed, this one has a structure filled with holes whereas we would like the blocks to be full. It is then possible to apply a closing to this new binary image, that fills the holes (smaller than the structuring element) while keeping the structure of the antidiagonal. This method will be systematically applied to the



(A) Opening of the self-similarity matrix

(B) Extraction of the blocks around the antidiagonal

(C) Closing of the blocks around the antidiagonal

FIGURE 4.10: Filtration of the self-similarity matrix by a 6x6 size structuring element.

antidiagonal to obtain a clearer image. The result of the closing by a structuring element (of size 100x100 always homogeneous equal to 1) is presented in Figure 4.10c. From this last image it is possible to detect the global structure of the piece. The dif-

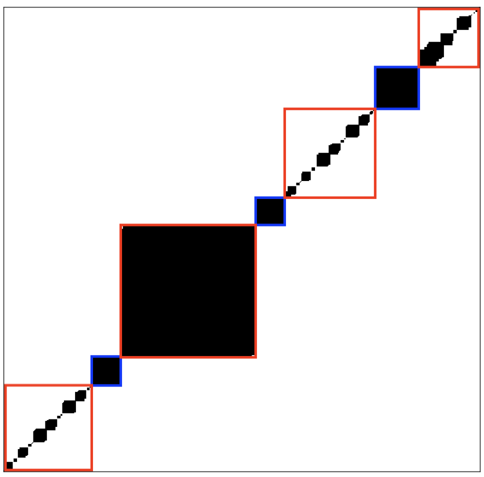
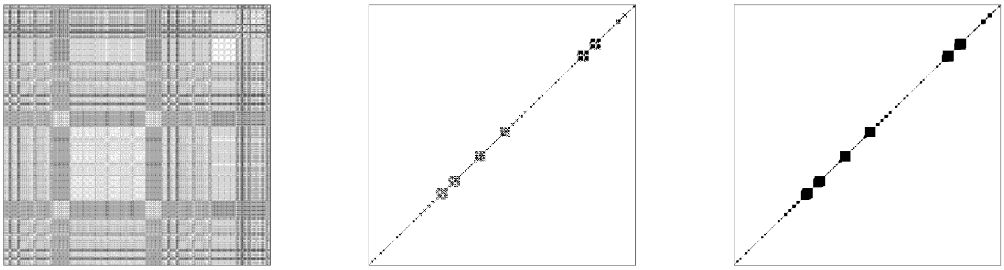


FIGURE 4.11: First detection of the different blocks around the antidiagonal to identify the global structure of the piece.

ferent detection blocks are shown in Figure 4.11. The red blocks represent the grey rectangles in Figure 4.8 while the blue blocks represent the C and C' passages. This method then works correctly to detect the global structure of the piece but it is possible to go further by detecting the different themes within the different blocks. To do this, we filter the self-similarity matrix with a smaller structuring element to detect smaller passages. By taking as structuring element a square of side 3 whose all coefficients are equal to 1, the opening of the self-similarity matrix becomes less clear but has more details which allow us to detect smaller blocks. The result is presented in Figure 4.12a. As before, the flood fill algorithm allows to extract the antidiagonal of this last matrix, which is represented in Figure 4.12b. As previously it is possible to apply a closing to make the blocks simply connected. This result is shown in Figure 4.12c. This last image allows us to detect the structure inside each block, i.e. to detect patterns of a duration similar to a bar. Since the global structure represented by the



(A) Opening of the self-similarity matrix (B) Extraction of the blocks around the antidiagonal (C) Closing of the blocks around the antidiagonal

FIGURE 4.12: Filtration of the self-similarity matrix by a 3x3 size structuring element.

red and blue blocks has been detected, all that remains is to look separately at the structure inside each block. By focusing on each block independently of the others, it is then possible to detect an internal structure within each block. It is possible to overlay the information obtained on each block with the more global information in Figure 4.11. These data on the structure of the piece are represented in Figure 4.13.

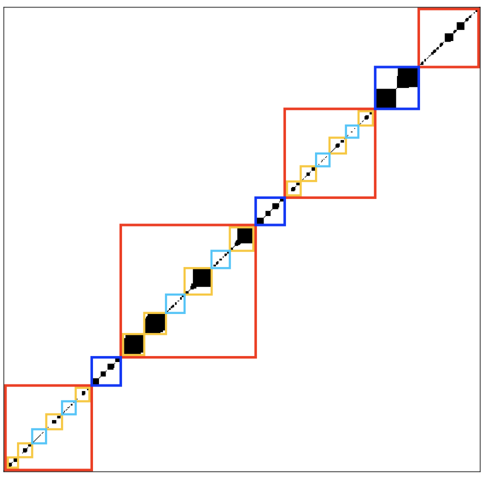
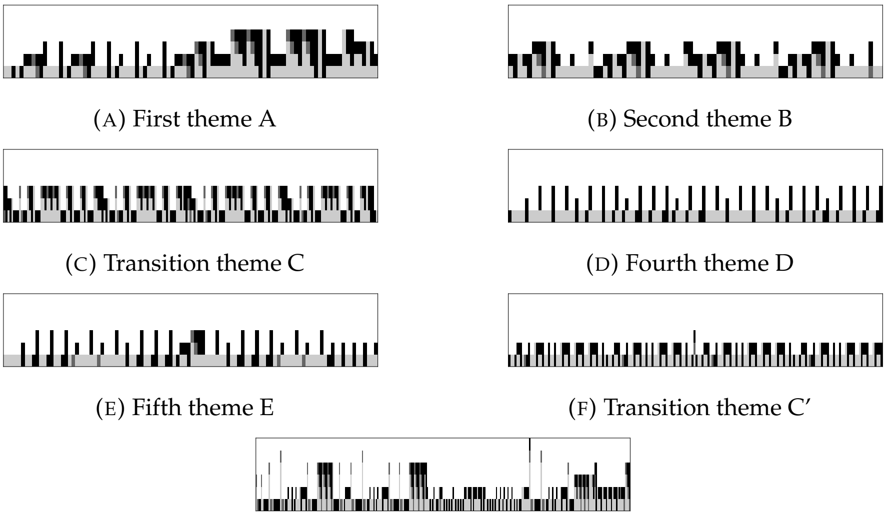


FIGURE 4.13: Representation of the different degrees of filtering which allows to observe a global and internal structure in each block.

Inside the first and third red blocks, a sub-structure emerges, the orange and light blue squares symbolize respectively the themes *A* and *B* in Figure 4.8. This phenomenon is repeated with the second red block, the orange and light blue squares represent the *D* and *E* themes. This confirms the whole structure of the piece that was in Figure 4.8. Once the structure of the piece has been found, it is possible to return to the chord contour to find the different melodic patterns. The different patterns in the chord contour identified with the figure 4.13 are represented in Figure 4.14.



(A) First theme A

(B) Second theme B

(C) Transition theme C

(D) Fourth theme D

(E) Fifth theme E

(F) Transition theme C′

(G) Final theme F

FIGURE 4.14: Different musical themes in Alla Turca.

Before finishing reading this chapter, you are invited to visualize the filtration of the self-similarity matrix of Alla Turca with an increasing structuring element. Video that is available here allows to observe the progression of the structure of the piece. Filtration begins with a 2x2 size structuring element and gradually increases to a 19x19 size structuring element. This makes it possible to reveal how the main musical themes are constructed.

Conclusion and Future Work

This report proposes to apply mathematical morphology to music in two different ways: binary mathematical morphology to a piano roll representation and grayscale mathematical morphology to a self-similarity matrix. First of all, Chapter 2 presents some direct applications of mathematical morphology to music by working on piano rolls. Morphological tools allow to analyze a musical piece such as isolating the melody or the chords and extracting the tonality. The results we obtained allow us to get a first view of what morphology can bring to music. Then, Chapter 3 proposes a new method to visualize the regularity of a piece of music: observe the melodic patterns. These patterns are based on the pitch variations along the piece, also called melodic contour, i.e. the normalized slope of the melodic curve to summarize the information. The originality of this approach comes from the fact that this idea is generalized to chords. If chords follow one another instead of notes, it will no longer be a number that represents the pitch variation but a matrix, called transition matrix that will characterize the sequence between two chords. Thus, melodic patterns become much richer, more varied and more complex. To detect them, Chapter 4 suggests working on the self-similarity matrix of the chord contour. With morphological tools, it is possible to have different degrees of filtering on this matrix. This provides information on the structure of the piece at different scales. Filtering the matrix with a fairly high degree of loss will highlight the main parts of the song, while filtering more precisely will make it possible to identify the different musical themes within each part. Therefore, the self-similarity matrix makes it possible to determine the structure of the piece and, by working with the chord contours, to find the melodic patterns that characterize the different musical themes.

In future work it would be possible to work on a similarity matrix between two chord contour from two different pieces. This would make it possible to identify common patterns and determine the proximity between these two pieces. Another way to explore would be to establish a classification of melodic patterns, to see if patterns are more present in some musical genres and in certain countries or cultures.

On a more personal point of view, this internship has allowed me to discover new methods of work, to develop my skills in musical analysis and to deepen my knowledge of mathematics in music. I am very satisfied to have been able to develop the idea presented in the last three chapters. This work clearly confirms my choice to do a PhD in musical computational analysis. Finally, during these six months spent at Ircam as an intern, I constantly interacted with many Ircam researchers and musicians through meetings at conferences organized or even around the coffee machine and during the lunch break. These meetings are very rich and allow us to understand the other research themes studied at Ircam, thus allowing us to play a part in the Ircam's ideology which is to gather researchers and musicians to work as a team on a shared project.

Appendix A

Music Pieces Analysis with the Pitch Variations Method

A.1 Analysis of the First Gymnopédie

The *Gymnopédies* are three works for piano composed by Erik Satie, published in Paris in 1888. We will focus on the first Gymnopédie. The chord contour, represented in Figure A.1, show that the structure of the song repeats itself twice. The following Figures illustrate the method of analysis proposed in Chapters 3 and 4.

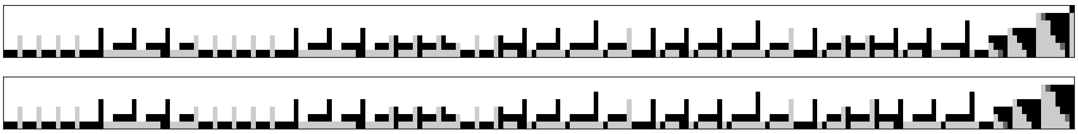


FIGURE A.1: Chord contour of the first Gymnopédie.

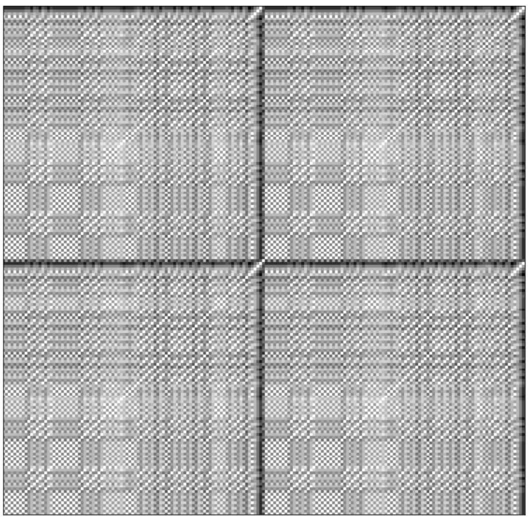
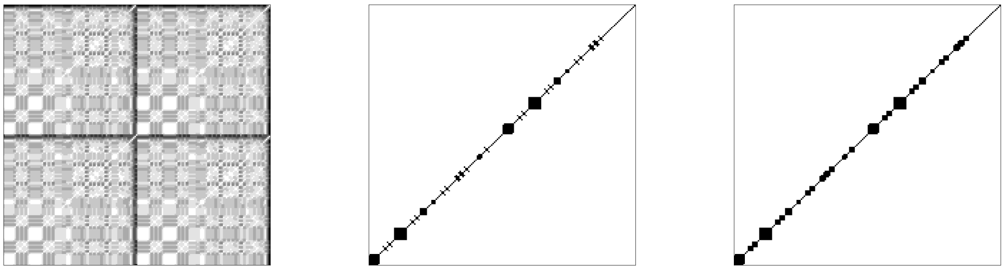
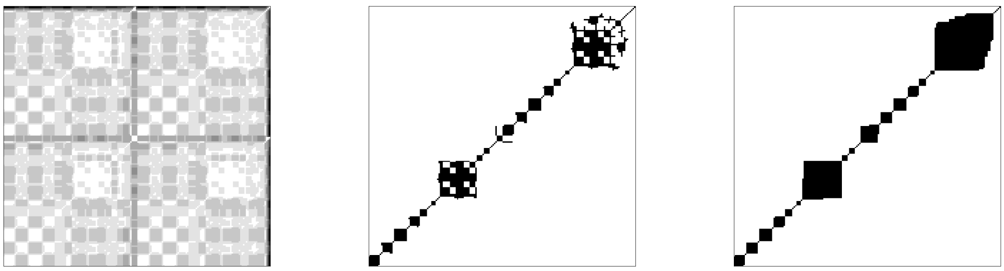


FIGURE A.2: Self-similarity matrix of the first Gymnopédie.



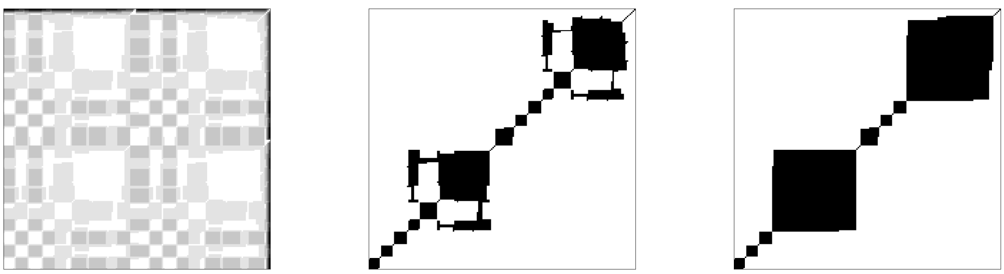
(A) Grayscale opening (B) Extraction of the blocks (C) Binary closing

FIGURE A.3: Filtering of the self-similarity matrix by a 2x2 size structuring element.



(A) Grayscale opening (B) Extraction of the blocks (C) Binary closing

FIGURE A.4: Filtering of the self-similarity matrix by a 4x4 size structuring element.



(A) Grayscale opening (B) Extraction of the blocks (C) Binary closing

FIGURE A.5: Filtering of the self-similarity matrix by a 8x8 size structuring element.

A.2 Analysis of the third movement of Moonlight Sonata

It is possible to study longer and more complex pieces like the third movement of the *Piano Sonata No. 14* by Ludwig van Beethoven. More popularly known as the third movement of *Moonlight Sonata*, the sonata of two hundred bars that evokes a funeral march was composed in 1801. The size of the piece results in a chord contour of a consequent length. The series was cut in eight parts for more clarity and represented in Figure A.6. As before, the other figures illustrate the method proposed in this report to determine the structure of the song.

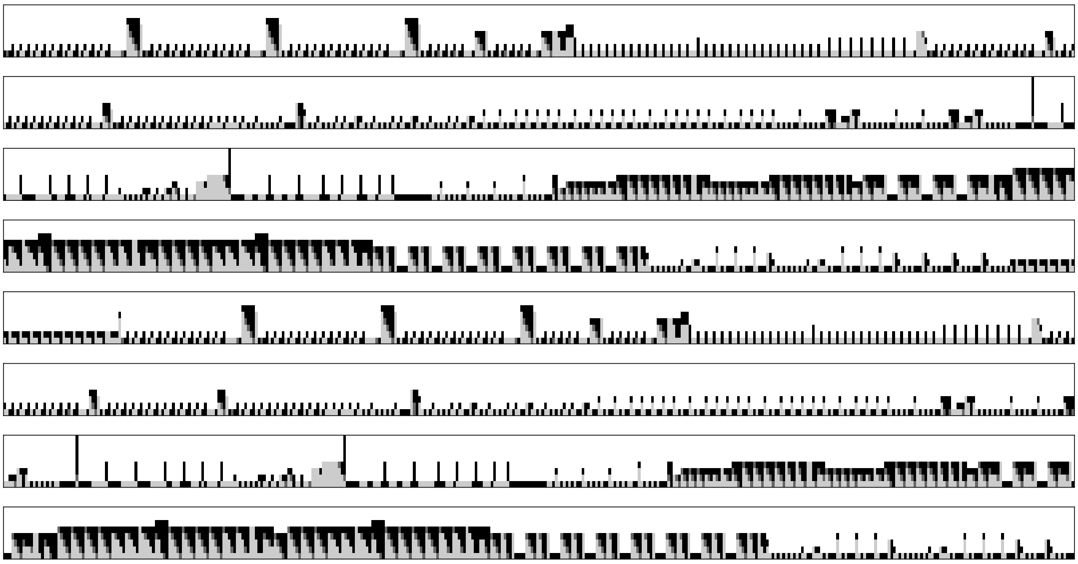


FIGURE A.6: Chord contour of the third movement of Moonlight Sonata.

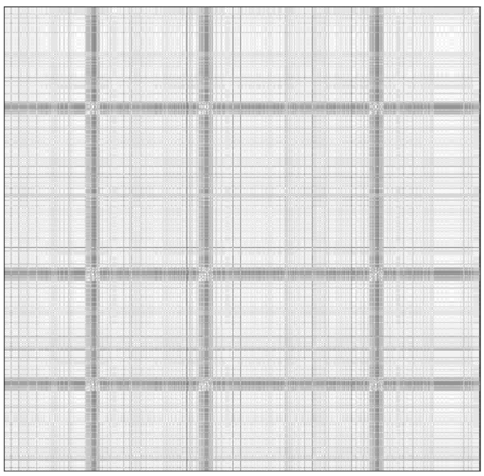
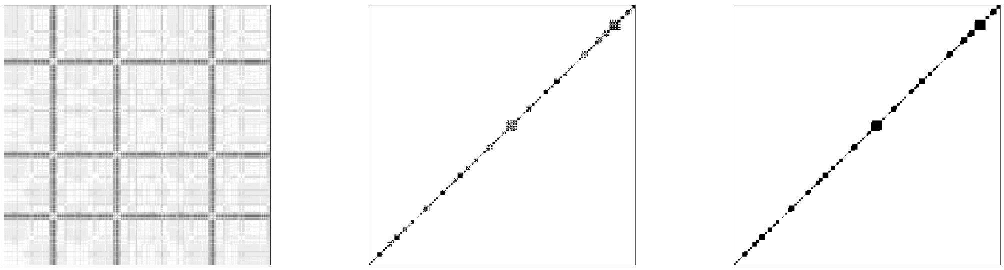
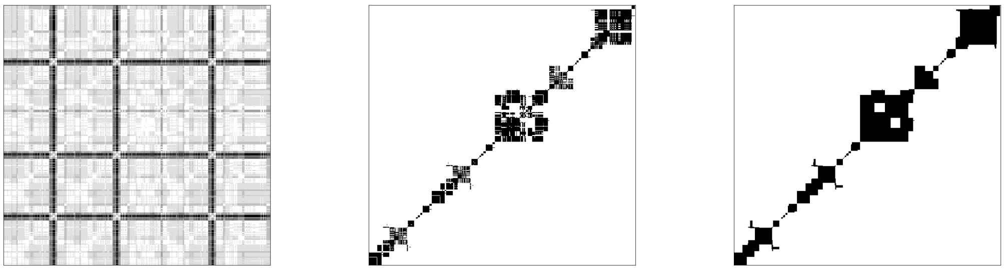


FIGURE A.7: Self-similarity matrix of the third movement of Moonlight Sonata.



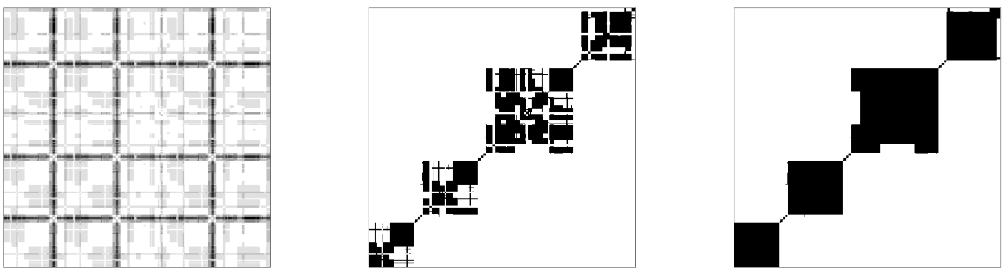
(A) Grayscale opening (B) Extraction of the blocks (C) Binary closing

FIGURE A.8: Filtering of the self-similarity matrix by a 4x4 size structuring element.



(A) Grayscale opening (B) Extraction of the blocks (C) Binary closing

FIGURE A.9: Filtering of the self-similarity matrix by a 5x5 size structuring element.



(A) Grayscale opening (B) Extraction of the blocks (C) Binary closing

FIGURE A.10: Filtering of the self-similarity matrix by a 17x17 size structuring element.

Bibliography

[1] I. Bloch, H. Heijmans, and C. Ronse, "Mathematical morphology," in *Handbook of Spatial Logics*, pp. 857–944, Springer Netherlands, 2007.

[2] J. Serra, *Image Analysis and Mathematical Morphology*. London: Academic Press, Inc., 1982.

[3] H. Heijmans, *Morphological Image Operators*. Advances in Electronics and Electron Physics: Supplements, Academic Press, 1994.

[4] H. Heijmans and C. Ronse, "The algebraic basis of mathematical morphology I. dilations and erosions," *Computer Vision, Graphics, and Image Processing*, vol. 50, no. 3, pp. 245 – 295, 1990.

[5] C. Ronse and H. Heijmans, "The algebraic basis of mathematical morphology: II. openings and closings," *CVGIP: Image Understanding*, vol. 54, no. 1, pp. 74 – 97, 1991.

[6] L. Najman and H. Talbot, *Mathematical Morphology: From Theory to Applications*. ISTE-Wiley, 2010.

[7] H. Minkowski, "Volumen und Oberfläche," *Mathematische Annalen*, vol. 57, pp. 447–495, 1903.

[8] H. Hadwiger, "Minkowskische Addition und Subtraktion beliebiger Punktmengen und die Theoreme von Erhard Schmidt," *Mathematische Zeitschrift*, vol. 53, pp. 210–218, 1950.

[9] M. Karvonen and K. Lemström, "Using mathematical morphology for geometric music information retrieval," in *International Workshop on Machine Learning and Music, MML'08, Helsinki, Finland*, 2008.

[10] M. Karvonen, M. Laitinen, K. Lemström, and J. Vikman, "Error-tolerant content-based music-retrieval with mathematical morphology," in *International Symposium on Computer Music Modeling and Retrieval*, vol. LNCS 6684, pp. 321–337, 2010.

[11] C. Agon, M. Andreatta, J. Atif, I. Bloch, and P. Mascarade, "Musical descriptions based on formal concept analysis and mathematical morphology," in *International Conference on Conceptual Structures*, pp. 105–119, Springer, 2018.

[12] S. E. Trehub, D. Bull, and L. A. Thorpe, "Infants' perception of melodies: The role of melodic contour," *Child Development*, vol. 55, no. 3, pp. 821–830, 1984.

[13] W. J. Dowling, "Melodic contour in hearing and remembering melodies," in *Musical Perceptions* (R. Aiello and J. A. Sloboda, eds.), p. 173–190, Oxford University Press, 1994.

[14] C. R. Adams, "Melodic contour typology," *Ethnomusicology*, vol. 20, no. 2, pp. 179–215, 1976.

[15] C. Buteau and G. Mazzola, "Motivic analysis according to Rudolph Réti: formalization by a topological model," *Journal of Mathematics and Music*, vol. 2, no. 3, pp. 117–134, 2008.

[16] C. Anagnostopoulou, M. Giraud, and N. Poulakis, "Melodic contour representations in the analysis of children's songs," in *3rd International Workshop on Folk Music Analysis* (P. van Kranenburg, C. Anagnostopoulou, and A. Volk, eds.), (Amsterdam, Netherlands), pp. 40–43, 2013.

[17] I. Quinn, "The combinatorial model of pitch contour," *Music Perception: An Interdisciplinary Journal*, vol. 16, no. 4, pp. 439–456, 1999.

[18] E. W. Marvin and P. A. Laprade, "Relating musical contours: Extensions of a theory for contour," *Journal of Music Theory*, vol. 31, no. 2, pp. 225–267, 1987.

[19] J. Foote, "Visualizing music and audio using self-similarity," in *Seventh ACM International Conference on Multimedia (Part 1)*, pp. 77–80, 1999.

[20] F. Lerdahl and R. S. Jackendoff, *A Generative Theory of Tonal Music, reissue, with a new preface*. MIT press, 1996.