Polytopic reconfiguration: a graph-based scheme for the multiscale transformation of music segments and its perceptual assessment
Valentin Gillot, Frédéric Bimbot

To cite this version:
Polytopic reconfiguration: a graph-based scheme for the multiscale transformation of music segments and its perceptual assessment

Valentin GILLOT  
IRISA / INSA, Rennes (France)  
valentin.gillot@irisa.fr

Frédéric BIMBOT  
IRISA / CNRS UMR-6074, Rennes (France)  
frederic.bimbot@irisa.fr

ABSTRACT

Music is a sequential process for which relations between adjacent elements play an important role. Expectation processes based on alternations of similarity and novelty contribute to the structure of the musical flow. In this work, we explore a polytopic representation of music, which accounts for expectation systems developing at several timescales in parallel. After recalling properties of polytopic representations for describing multi-scale implication processes, we introduce a scheme for recomposing musical sequences by simple transformations of their support polytope. A specific set of permutations (referred to as Primer Preserving Permutations or PPP) are of particular interest, as they preserve systems of analogical implications within musical segments. By means of a perceptual test, we study the impact of PPP-based transformations by applying them to the choruses of pop songs in midi format and comparing the result with Randomly Generated Permutations (RGP). In our test, subjects are asked to rate musical excerpts reconfigured by PPP-based transformations versus RGP-based ones in terms of musical consistency and of attractiveness. Results indicate that PPP-transformed segments score distinctly better than RGP-transformed for the two criteria, suggesting that the preservation of implication systems plays an important role in the subjective acceptability of the transformation. Additionally, from the perspective of building an automatic recomposition system for artistic creation purposes, we introduce, in appendix, the preliminary version of an automatic method for decomposing segments into low-scale musical elements, taking into account possible phase-shifts between the musical surface of the melody and the metrical information.

1. INTRODUCTION

Music is usually considered as a sequential process, where sounds, group of sounds and motifs are occurring chronologically, following the natural unfolding of time. Under this approach, music is considered as a flow of information, the organization of which is essentially governed by the relations that develop between adjacent elements. This property of sequentiality is indeed central to many models of music description and generation, including musicological models such as Schenkerian analysis [1], generative theories based on tree-like structures such as GTTM [2], or computational automata such as Markov chains [3].

From a complementary perspective, it is also commonly accepted that the organization of music is widely based on alternations of similarity and novelty which create patterns and systems of expectation and surprise that ultimately contribute to the structure of the musical discourse from a cognitive point of view [4] [5] [6]. These can be represented as a Polytopic Graph of Latent Relations [7] where each node of the graph represents a low-scale musical segment and vertices correspond to their relation within the expectation systems.

The aim of this work is to explore the polytopic model in terms of its relevance to account for the inner structure of musical segments, by assessing its potential use for the structural recomposition of music. In fact, as presented in section 2, the polytopic representation of a musical segment enables a large range of transformations by applying various permutations to its nodes, thus generating multiple reconfigurations of its musical content, with the same elements in a different order.

Specific permutations, called Primer Preserving Permutations (PPP), are of particular interest, as they preserve systems of analogical implications between metrically homologous elements within the segment [8]. The central hypothesis of the present work is that the musical consistency of PPP-transformed segments will therefore be less affected than it would be by an “ordinary” (i.e. randomly generated) permutation.

In section 2, we describe the implementation of the polytopic reconfiguration process and we elaborate on the organizational properties of Primer Preserving Permutations as well as their potential impact on the inner structure of musical segments.

Then, in section 3, we assess the relevance of the reconfiguration scheme (and its underlying hypotheses): we report on a perceptual test where subjects are asked to rate musical properties of MIDI segments, some of them have been reconfigured with PPPs while others were transformed by Randomly Generated Permutations (RGP) designed so as to possess a comparable number of discontinuities.

In appendix, we introduce an automatic method for decomposing segments into low-scale musical elements, taking into account possible phase-shifts between the musical surface and the metrical information (for instance, anacruses).
2. BACKGROUND AND FORMALISM

2.1 Musical objects and time-scales

Before detailing the background and formalism aspects of this work, it is important to specify what musical objects are under consideration in our study. Following Snyder’s “levels of musical experience” [9], one may identify three ranges of time-scales in music: an event-fusion level, up to 1/32<sup>th</sup> second (corresponding to early processing), an intermediate “melodic and rhythmic grouping” level, between 1/16<sup>th</sup> and 8 sec. (governed by short-term memory) and a form level, 16 sec. or above (resorting to long-term memory).

In this work, we focus on sectional units (i.e. the first time-scale of the form level) and we study their inner organization as the result of the relations existing between successive constituents at the intermediate level, around 1 sec (i.e. typically half-bars). We put a particular focus on melody, with the consequence that our low-scale objects are small groups of a few notes with a common harmonic background.

2.2 Polytopic representations of implication systems

The basic concept of a sequential implication system is that the observation of a given event by a subject with no particular preconception will trigger (to that subject) the expectation of the next event is likely to be similar to the particular preconception. This is considered as even more so, in the case of expectation that the next event is likely to be similar to the observation of a given event by a subject with no particular preconception will trigger (to that subject) the expectation of the next event is likely to be similar to the particular preconception.

In this work, we focus on sectional units (i.e. the first time-scale of the form level) and we study their inner organization as the result of the relations existing between successive constituents at the intermediate level, around 1 sec (i.e. typically half-bars). We put a particular focus on melody, with the consequence that our low-scale objects are small groups of a few notes with a common harmonic background.

2.3 Structural reconfiguration by node permutations

There are many ways to define a permutation, but the most adequate one in the context of this work is to view it as a bijective function \( \varphi \) (i.e. a one-to-one correspondence) between the \( N \) time indexes of an original sequence \( x_0...x_{N-1} \) and those of the permuted sequence \( x_{\varphi(0)}...x_{\varphi(N-1)} \).

Applying a permutation to the time indexes creates a certain degree of disruption in the original flow of the sequence. In this paper, we characterize this fact by two properties: \( D \), the number of discontinuities created by the permutation and \( E \), the \( \log_2 \) (of the) maximum time-interval (or excursion) between 2 consecutive elements in the permuted sequence.

\[
D = \# \{ t \mid |\varphi(t) - \varphi(t-1)| \neq 1 \}_{0 < t < N},
\]
\[
E = \log_2 \max_{0 < t < N} |\varphi(t) - \varphi(t-1)|
\]

Among all possible permutations \( \varphi \), a particular subset of them is focused on in this work. They are referred to as PPPs (for Primer Preserving Permutations) and were introduced by Louboutin et al. [8] as the set of permutations which preserve the systemic relations of the musical elements \( x_t \) in the sequence, by just interchanging the time-scales at which they develop. These permutations are therefore particularly interesting to investigate on the relevance of the implications systems (and their preservation) in the consistency of a musical segment.

In the case of a sequence of \( N=16 \) elements, the polytope is a 4-cube, also called a tesseract (see Figure 2), and it can be viewed as being composed of 4 homologous lower-scale systems of 4 elements (formed by 4 parallel faces of the tesseract), themselves linked by an upper-scale system.
formed by the first element of each lower-scale system (on a face which is perpendicular to the 4 lower-scale ones) (see Figure 3).

Mapping these systems of faces by a permutation, while preserving time order within faces, yields 5 possibilities (plus the identity), which correspond to the PPPs defined in Table 1. Note that each PPP maps 0 to 0 and N-1 to N-1.

By construction, PPPs preserve implication systems of 4 elements in the sense that sets of 4 parallel faces in the tesseract are mapped on another set of parallel faces. By doing so, analogical systems are preserved, and only the scales at which they develop are modified: for instance, \([x_0x_1x_2x_3]\) in the original is “transferred” to \([x_0x_1x_4x_5]\) with PPP1 and \([x_0x_4x_8x_{12}]\) with PPP4. Moreover, the last element of a system never occurs before all the elements of all the systems it belongs to, have themselves occurred, therefore preserving a principle of causality in all the analogical implications.

For comparison purposes, Table 2 illustrates 5 Randomly Generated Permutations (RGPs), which also start with 0, end with N-1, and have the same profile as their corresponding PPPs in terms of discontinuity and excursion. It can be easily checked that, unless incidentally, analogical systems are not preserved by RGPs.

Ultimately, a given permutation is applied to a musical sequence of events by reading and copying (in the permuted order) the original musical material after having segmented into proper low-scale elements. Figure 4 illustrates this process on a 8-bar melodic line (original at the top), processed as a sequence of \(N=16\) half-bar (2-beat) elements, resulting in reconfigured melodic lines, the organization of which is either driven by PPPs or by RGPs.

---

**Table 1.** Extensive definition of the 6 PPPs and their properties (discontinuity and excursion - see text).

<table>
<thead>
<tr>
<th>PPP</th>
<th>ORIG</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>Disc.</th>
<th>Exc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>14</td>
<td>9</td>
<td>11</td>
<td>13</td>
<td>15</td>
<td>14</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>10</td>
<td>11</td>
<td>4</td>
<td>5</td>
<td>12</td>
<td>13</td>
<td>6</td>
<td>7</td>
<td>14</td>
<td>15</td>
<td>7</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>12</td>
<td>1</td>
<td>5</td>
<td>9</td>
<td>13</td>
<td>2</td>
<td>6</td>
<td>10</td>
<td>14</td>
<td>3</td>
<td>7</td>
<td>11</td>
<td>15</td>
<td>15</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>11</td>
<td>4</td>
<td>6</td>
<td>12</td>
<td>14</td>
<td>5</td>
<td>7</td>
<td>13</td>
<td>15</td>
<td>15</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.** Examples of RGPs and their properties (discontinuity and excursion - see text).

<table>
<thead>
<tr>
<th>RGP</th>
<th>1</th>
<th>0</th>
<th>3</th>
<th>4</th>
<th>2</th>
<th>7</th>
<th>1</th>
<th>6</th>
<th>5</th>
<th>8</th>
<th>13</th>
<th>10</th>
<th>9</th>
<th>11</th>
<th>14</th>
<th>12</th>
<th>15</th>
<th>Disc.</th>
<th>Exc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>2</td>
<td>7</td>
<td>15</td>
<td>14</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>14</td>
<td>15</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>14</td>
<td>2</td>
<td>10</td>
<td>1</td>
<td>9</td>
<td>12</td>
<td>7</td>
<td>11</td>
<td>13</td>
<td>15</td>
<td>15</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>9</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>10</td>
<td>3</td>
<td>11</td>
<td>14</td>
<td>13</td>
<td>5</td>
<td>12</td>
<td>15</td>
<td>15</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Example of the polytopic representation of a time sequence in the case of a tesseract.

By construction, PPPs preserve implication systems of 4 elements in the sense that sets of 4 parallel faces in the tesseract are mapped on another set of parallel faces. By doing so, analogical systems are preserved, and only the scales at which they develop are modified: for instance, \([x_0x_1x_2x_3]\) in the original is “transferred” to \([x_0x_1x_4x_5]\) with PPP1 and \([x_0x_4x_8x_{12}]\) with PPP4. Moreover, the last element of a system never occurs before all the elements of all the systems it belongs to, have themselves occurred, therefore preserving a principle of causality in all the analogical implications.

For comparison purposes, Table 2 illustrates 5 Randomly Generated Permutations (RGPs), which also start with 0, end with \(N-1\), and have the same profile as their corresponding PPPs in terms of discontinuity and excursion. It can be easily checked that, unless incidentally, analogical systems are not preserved by RGPs.

Ultimately, a given permutation is applied to a musical sequence of events by reading and copying (in the permuted order) the original musical material after having segmented into proper low-scale elements. Figure 4 illustrates this process on a 8-bar melodic line (original at the top), processed as a sequence of \(N=16\) half-bar (2-beat) elements, resulting in reconfigured melodic lines, the organization of which is either driven by PPPs or by RGPs.

2.4 Handling musical surface time-shifts

As a prerequisite to the application of a permutation to a sequence of musical objects, it is of primary importance to define precisely the location and boundaries of the objects that will be subjected to the transformation.

In this work, elementary objects are melodic fragments
of 1/2-bar long which correspond to 2 beats in the case of a 4/4 meter (as is the case in the example of Fig 4).

However, for reconfiguration purposes, these “objects” may not be optimal if they are blindly synchronized on downbeats (and sub-downbeats). In a number of cases, the musical surface of the “proper” low-scale melodic elements should be slightly shifted in order to correspond to an optimal grouping of the notes and therefore a better result after permutation.

This situation is illustrated in Figure 5 where the first melodic element should be considered as starting 1.5 beat before the first downbeat of the section.

In designing the perceptual test described in the next section, this optimal time-shift has been adjusted manually on a case-by-case basis. However, we present in the appendix, a preliminary algorithm that has been developed to estimate the time-shift automatically.

3. PERCEPTUAL EXPERIMENTS IN STRUCTURAL RECONFIGURATION

In this section, we present the details of the methodology which has been used for the evaluation of polytopic reconfigurations. The general principle has been to design a perceptual test where subjects are presented with musical excerpts which are either PPP-transformed or RGP-transformed, and where they are asked to rate (blindly) these excerpts both in terms of consistency and attractiveness. We then study the difference of ratings across subjects between the two types of transformations.

3.1 Test data

The experiments are performed on the RWC POP corpus, a subset of the Real World Computing (RWC) music database created for scientific purposes by Goto et al. [14]. This dataset has been designed for researchers. It is available for a small fee and without restrictions.

The RWC POP set is made of 100 songs in WAV and MIDI formats. According to the authors, they were composed partly in the style of the 80s American hits and partly in the 90s Japanese style. This corpus is very commonly used in various task in Music Processing and Information Retrieval.

Using the RWC songs in our tests has an additional advantage: there is virtually no risk that subjects participating to the test already know the songs from which the excerpts are stemming. This is a favorable situation as the familiarity with a particular song could affect the rating of the transformed excerpts in an uncontrolled way.

For our experiments, a subset of 24 songs has been selected for which the first instance of chorus is exactly 8-bar long. The reconfiguration process is applied to the MIDI version of the chorus. The elements of interest are the melodic lines, the accompaniment (harmony and bass) and the drums. We consider that these elements provide an acceptable rendering of the structural information of the
the possibility to position the cursor between two gradu-
thors, selected among colleagues and relatives of the au-
As presented in section 2, these transformations are per-
formed on the original chorus after having segmented it into 16 elements (with a possible shift of the musical sur-
face of the melody, as explained in section 2.4). Then, the 10 permutations (5 PPPs and 5 RGPs) are applied to the 16 elements, resulting in a re-ordering of the original chor-
uses, i.e. 10 new musical contents (per chorus), which we will call “variants” of the original.
To summarize, for each of the 24 choruses, the following 11 variants are considered:
• the original version of the chorus
• 5 PPP-transformed variants (PPP₁ to PPP₅)
• 5 RGP-transformed variants (RGP₁ to RGP₅)
The perceptual test is therefore conducted on 24 × 11 =
66 distinct excerpts.

3.2 Protocol
The perceptual experiments are performed with 66 sub-
jects, selected among colleagues and relatives of the au-
1. “Please give your opinion as of the consistency of
the musical construction of this passage”.
2. “Please give your opinion as of the degree of musical
attractiveness of this passage”
The subjects must respond by placing a cursor on a 5-
level scale corresponding to 5 different levels of evaluation
(from very bad consistency to very good consistency and
from very unattractive to very attractive). They also have
the possibility to position the cursor between two gradu-
tions, thus offering them 9 levels of rating altogether.
Asking two different questions aims at leading subjects
to decorrelate as much as possible their rating of the ac-
ceptability of the musical construction (question 1), from
their own personal taste (which they can express in their
response to question 2).

3.3 Scoring
For each subject, we thus collect 24 consistency ratings
and 24 attractiveness ratings.
Let us denote as r, the generic variable which designates
such ratings (ignoring for the moment whether it is consis-
tency or attractiveness, for the sake of notations’ simplic-
ity).
The 24 ratings rjk produced by subject j (1 ≤ j ≤ 66,
1 ≤ k ≤ 24) are distributed into the 11 variants: O for or-
iginal, Pᵢ for PPPᵢ (1 ≤ i ≤ 5) and Pᵢ for RGPᵢ (1 ≤ i ≤ 5). These individual ratings are averaged separately,
thus leading to 11 mean ratings (one per type of variant):

\[
\mu_{0j} = \frac{1}{m_{0j}} \sum_{k=1}^{24} r_{jk} \quad \text{for the original choruses} \quad (1)
\]
\[
\mu_{ij} = \frac{1}{m_{ij}} \sum_{k=1}^{24} p_{i} r_{jk} \quad \text{for the 5 PPP variants} \quad (2)
\]
\[
\mu_{ij}^* = \frac{1}{m_{ij}} \sum_{k=1}^{24} p_{i}^* r_{jk} \quad \text{for the 5 RGP variants} \quad (3)
\]
Note that, in practice, m₀, mᵢ and mᵢ* are equal to 2 or 3 and their sum is equal to 24.
To evaluate the relative degradation created by the various
reconfigurations of the music material, we can measure:

\[
\delta_{ij} = \mu_{0j} - \mu_{ij} \quad (4)
\]
\[
\delta_{ij}^* = \mu_{0j} - \mu_{ij}^* \quad (5)
\]
namely the signed difference, for a given subject j, be-
tween his/her ratings of the reconfigured variants and his/her
ratings of the originals.
For each subject and variant, we can also compute:

\[
\pi_{ij} = \mu_{ij} - \mu_{ij}^* = \delta_{ij} - \delta_{ij}^* \quad (6)
\]
which measures the (positive or negative) preference of
PPPᵢ-transformed excerpts over RGPᵢ-transformed ones,
for speaker j (where PPPᵢ and RGPᵢ are comparable in
terms of discontinuities D and excursion E).
Ultimately, we focus on 3 scores:

\[
\pi_{j} = \frac{1}{5} \sum_{i=1}^{5} \pi_{ij} \quad (7)
\]
\[
\delta_{i} = \frac{1}{66} \sum_{j=1}^{66} \delta_{ij} \quad (8)
\]
\[
\delta_{i}^* = \frac{1}{66} \sum_{j=1}^{66} \delta_{ij}^* \quad (9)
\]

Figure 5. Example of a melodic line starting 1.5 beat before the first downbeat of the section.

chorus and they all are transformed synchronously by the
permutations. The average duration of a chorus is 19.5 sec-
onds.
In order to avoid the bias that could arise if a subject gets
familiar with a particular chorus, each subject is presented
with one and only one variant of each of the 24 choruses,
in a random order. The total number of tests in our ex-
periments is therefore equal to 66 × 24 = 1584, i.e. 144
tests per variant. Moreover, the 11 variants are distributed
as evenly as possible for each subject, i.e. typically 2 (or
occasionally 3) per subject.
After a brief general presentation of the test (during which
the subjects are not provided with any detail on the sci-
entific motivation of the test), they are asked to answer, for
each excerpt, two questions formulated as follows:

1. “Please give your opinion as of the consistency of

2. “Please give your opinion as of the degree of musical

attractiveness of this passage”

For obvious reasons, the authors themselves were not included in the panel of subjects.
where \( \pi_j \) provides a global preference score (positive or negative) on PPPs over RGPs for each subject \( j \), while \( \delta_i \) (resp. \( \delta_i^* \)) provides a degradation score for each of the PPP- (resp. RGP-) reconfigured variants, averaged over all subjects.

These two quantities are calculated separately for consistency ratings and for attractiveness ratings.

### 3.4 Results

Figure 6 and 7 depict the distributions of consistency and attractiveness preference scores \( (\pi_j) \), ranked from lowest to highest, over the panel of 66 subjects. In both cases, the distributions are visibly shifted towards positive values, especially for consistency ratings. Indeed, for consistency, 57.5 subjects have a positive preference for PPPs versus 8.5 having a negative preference \(^2\) (i.e. 87.1 % vs 12.9 %). The average score for positive judgments amounts to +0.89 whereas it rests at −0.30 only, for negative judgments. For attractiveness, the proportions are still clearly in favor of PPPs, but not as contrasts (74.2 % vs 25.8 %), with average scores of +0.74 and −0.55 respectively.

Figure 8 and 9 represent the relative degradation scores \( (\delta_i \text{ and } \delta_i^*) \) for the 10 variants, ranked in increasing order of degradation. From the ranking of the permutations, it is noticeable that the degradation score is primarily correlated with the number of discontinuities of the permutations, as the two permutations with lower \( D \) show a smaller value of \( \delta \) as opposed to the three others.

Globally, these results indicate that PPP-based reconfigurations tend to be less disruptive than RGPs against the perceived consistency of musical segments, thus supporting the hypothesis that the preservation of expectation systems contributes to their inner structure.

### 4. CONCLUSIONS

Systems of relations between elements play a key role in structural constructions in many cognitive dimensions. The experiments reported in this paper constitute an initial investigation on the role of analogical systems (modeled by polytopic graphs) in the perception of segmental structure in pop music.

These results indicate a very noticeable trend towards the hypothesis that the preservation of the analogical implication systems in musical segments impacts positively the perception of their inner structure and hence, their acceptability. However, more extensive tests on a larger population and with more varied musical excerpts are certainly needed to consolidate these first results, and refine their precise scope.

In parallel, we intend to develop the proposed reconfiguration scheme into a music creation tool which could be used by composers to experiment new possibilities in employing musical material with multiple purposes, results and effects. In this context, the polytopic representation is an additional advantage, as it is bound to make manipulation interfaces more intuitive and easy to use.

Ultimately, the development of an automatic segmentation process is a complementary goal towards a fully operational concept.

\(^2\) counting as 0.5 in each category, the subject who scores exactly 0.
Appendix

Melodic segmentation with a compressibility criterion

In order to perform a relevant permutation-based transformation to a sequence of musical objects, it is crucial to define precisely the location and boundaries of the objects on which the transformation is applied.

As mentioned in section 2.4, it is quite common that the structure of the melodic line is time-shifted with respect to the phase of the metrical pulsation. As a consequence, it is important to take into account this time-shift when defining the low-level melodic objects. In this section, we present a preliminary method to automatically estimate this time-shift.

For doing so, we estimate a segmentation of the melodic surface by optimizing the grouping of the notes using a compressibility criterion.

Meter, segmentation and phase-shift

For simplicity and consistency with the rest of our work, we consider musical segments which are 8-bar long, and 4/4 meter, i.e. \(L = 32\) beats, and we assume that we want to segment the excerpt in \(m\) elementary objects. In this work \(m = 16\).

Let \(S\) be the sequence of instants where the strong beats fall, according to the meter and the bar:

\[
S = (t_0, t_1, ..., t_p, ..., t_{m-1})
\]

Let us now denote as \(\Sigma\) another segmentation that is not necessarily synchronized with downbeats.

\[
\Sigma = (\tau_0, \tau_1, ..., \tau_p, ..., \tau_{m-1})
\]

In the general case we can write:

\[
\tau_p = t_p + \theta
\]

where \(\theta\) is the phase-shift of the musical surface of segment \(p\) in relation to the strong beat location \(t_p\).

In compliance with the work reported in the body of this article, we further assume that \(\theta\) does not depend on \(p\):

\[
\tau_p = t_p + \theta
\]

Properties of the phase-shift

When \(\theta = 0\), the melodic segments are considered to be starting synchronously with the downbeat. If \(\theta > 0\), this means that the melodic objects display some phase-delay with the beat, whereas if \(\theta < 0\), they start in anticipation (as in the case of an anacrusis).

Let \(\alpha\) be the “size” of one elementary object in terms of the number of beats, i.e.:

\[
\alpha = \frac{L}{m}
\]

Here \(\alpha = 2\).

Quite naturally, we assume that the phase-shift cannot exceed the size of the elementary objects, i.e.:

\[
-\alpha \leq \theta \leq \alpha
\]

For simplicity and consistency with the rest of our work, we consider musical segments which are 8-bar long, and 4/4 meter, i.e. \(L = 32\) beats, and we assume that we want to segment the excerpt in \(m\) elementary objects. In this work \(m = 16\).

Let \(S\) be the sequence of instants where the strong beats fall, according to the meter and the bar:

\[
S = (t_0, t_1, ..., t_p, ..., t_{m-1})
\]

For the time being, \(\lambda\) is tuned empirically.

We then define an elementary cross-compressibility score function \(c(\sigma_p, \sigma_q)\) between two melodic objects \(\sigma_p\) and \(\sigma_q\), by combining \(d\) and \(b\):

\[
c(\sigma_p, \sigma_q) = d(\sigma_p, \sigma_q) - \lambda b(\sigma_p, \sigma_q)
\]

For the time being, \(\lambda\) is tuned empirically.

We then define a compressibility score function for each elementary object \(\sigma_p\) within \(\Sigma\), which depends on the previous objects in the segmentation, for instance:

\[
z(\sigma_p) = \frac{1}{p} \sum_{h<p} c(\sigma_h, \sigma_p)
\]

Finally, we compute the entire segmentation compressibility score \(Z(\Sigma)\) as:

\[
Z(\Sigma) = \frac{1}{m} \sum_p z(\sigma_p)
\]
Because \( \theta \) is the same across segments, all possible segmentations \( \Sigma \) can be tested exhaustively, and the optimal segmentation is chosen as the one with the highest compressibility score.

Fig. 11 depicts the behavior of \( Z \) for a particular example, with \( \theta \) ranging from \(-2\) to \(2\) by steps of \(0.25\) beat. Score values are normalized to \(1\) for the maximum value.

**Figure 11.** Behavior of the compressibility score \( Z \) for a particular 8-bar melody, with \( \theta \) ranging from \(-2\) to \(2\) beats.

**Current performance level**

In its current version, the results provided by this algorithm have been evaluated over the 24 choruses used in the perceptual experiments, by comparing the automatic estimations of the melodic phase-shifts with those determined by a manual annotation: in 10 cases (out of 24), the automatic algorithm provided the same result as the expert.

This approach is currently being improved but the current results encourage us to further investigate the method, so as to evaluate its full potential.

5. REFERENCES


