Modeling the multiscale structure of chord sequences using polytopic graphs

Corentin Louboutin, Frédéric Bimbot

MODELING THE MULTISCALE STRUCTURE OF CHORD SEQUENCES USING POLYTOPIC GRAPHS

Corentin Louboutin
Université Rennes 1 / IRISA, France
corentin.louboutin@irisa.fr

Frédéric Bimbot
CNRS - UMR 6074 / IRISA, France
frederic.bimbot@irisa.fr

ABSTRACT

Chord sequences are an essential source of information in a number of MIR tasks. However, beyond the sequential nature of musical content, relations and dependencies within a music segment can be more efficiently modeled as a graph.

Polytopic Graphs have been recently introduced to model music structure so as to account for multiscale relationships between events located at metrically homologous instants.

In this paper, we focus on the description of chord sequences and we study a specific set of graph configurations, called Primer Preserving Permutations (PPP). For sequences of 16 chords, PPPs account for 6 different latent systems of relations, corresponding to 6 main structural patterns (Prototypical Carrier Sequences or PCS). Observed chord sequences can be viewed as distorted versions of these PCS and the corresponding optimal PPP is estimated by minimizing a description cost over the latent relations.

After presenting the main concepts of this approach, the article provides a detailed study of PPPs across a corpus of 727 chord sequences annotated from the RWC POP database (100 pop songs). Our results illustrate both qualitatively and quantitatively the potential of the proposed model for capturing long-term multiscale structure in musical data, which remains a challenge in computational music modeling and in Music Information Retrieval.

1. INTRODUCTION

One of the essential properties of music structure is the multiscale nature of the inner organization of musical segments, i.e. the existence of relationships between musical elements at different time-scales simultaneously.

Given its important role in supporting the local harmonic ground-plan of the music in a significant number of music genres, chord sequences are commonly considered as an essential source of information in a variety of MIR tasks (see for instance [13, 17, 22]).

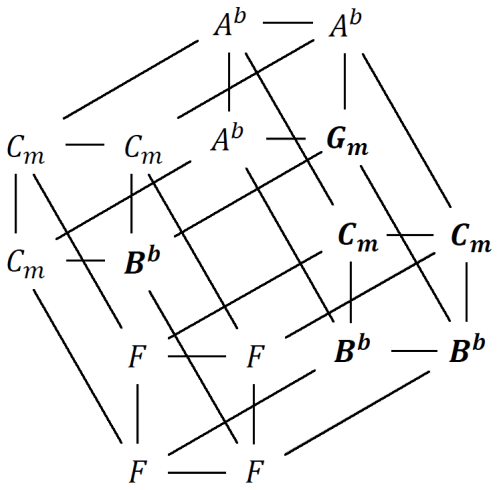

Figure 1. A chord sequence represented on a tesseract.

However, beyond the sequential order of chords along the timeline, relations and dependencies between chords within a music segment tend to be more efficiently modeled as a graph.

Polytopic Graphs of Latent Relations (PGLR) [11] have been recently introduced to model music structure, so as to account for multiscale relationships between events located at metrically homologous instants, by means of an oriented graph supported by a n-dimensional polytope.

This class of models is assumed to be particularly well suited for strongly "patterned" music, such as pop music, where recurrence and regularity tend to play a central part in the structure of the musical content.

PGLR also relax the adjacency hypothesis of the GTTM model [10], by which the grouping of elements into higher level objects is strictly limited to neighbouring units. This is particularly useful to account for period-like $abac$ patterns, where the similarity relationship between the two a segments spans above (and irrespective of) the b segment.

In this paper, we focus on the description of metric-synchronous chord sequences of 16 elements, resting on regular phrasal structures or *carrures*. In that case, the supporting polytope is a tesseract (i.e. a 4-cube) as illustrated by Fig. 1, and the graph description lives on this tesseract (as represented on Fig. 4).

After providing, in Section 2, the main concepts and formalisms related to the approach, we study in detail a particular variant of the model, where the graph structure

is restricted to a set of 6 configurations, called Primer Preserving Permutations (PPP). We show in Section 3.1 that PPPs relate to prototypical multi-scale structural patterns which we call Prototypical Carrier Sequences (PCS) and we explain how observed chord sequences can be viewed as distorted versions of these prototypical patterns. In the last part of the article (Section 4), we provide an experimental study of PPPs across a corpus of 727 chord sequences annotated from the RWC POP database (100 pop songs) with qualitative and quantitative results illustrating the potential of the model. We conclude with perspectives outlined by the proposed approached.

2. CONCEPTS AND FORMALISM

2.1 The PGLR Framework

As mentioned in the introduction, the PGLR approach views a sequence of musical elements within a structural segment as exhibiting privileged relationships with other elements located at similar metrical positions across different timescales.

For metrically regular segments of 2^n events, the corresponding PGLR conveniently lives on an n-dimensional cube (square, cube, tesseract, etc...) [1] , n being the number of scales considered simultaneously in the multiscale model. Each vertex in the polytope corresponds to a musical element of the lowest scale, each edge represents a latent relationship between two vertices and each face forms an elementary system of relationships between (typically) 4 elements. In addition, the proposed model views the last vertex in each elementary system as the denied realization of a (virtual) expected element, itself resulting from the implication triggered by the combination of former relationships within the system (see Section 2.3).

For a given support polytope, the estimated PGLR structure results from the joint estimation of (i) the configuration of an oriented graph resting on the polytope, with the constraint that it reflects causal time-order preserving dependencies and interactions between the elements within the musical segment, and (ii) the inference of the corresponding relations between the nodes of the graph, these relations being termed as latent, as they are not explicitly observed (and may even not be uniquely defined).

2.2 Chord Representation and Relations

Strictly speaking, a chord is defined as any harmonic set of pitches that are heard as if sounding simultaneously. However, in tonal western music, chords are more specifically understood as sets of *pitch classes* which play a strong role in the accompaniment of the melody (in particular, in pop songs).

A number of formalisms exist for describing chord relations, either in the context of classical musicology or in the framework of more recent theories, for instance, the neo-Riemannian theory and voice-leading models [5, 6, 20], or computational criteria such as Minimal Transport [10].

[1] and more generally speaking, on an n-polytope

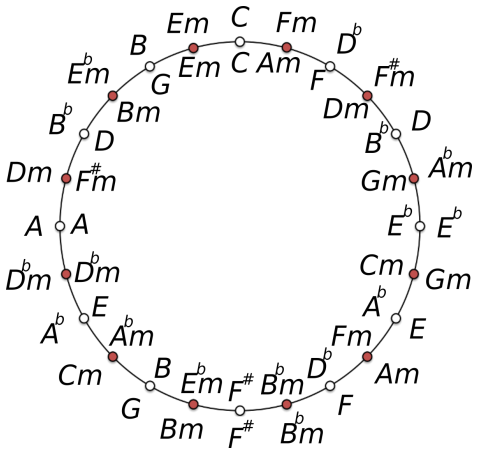

Figure 2. Circles of thirds (inner) and phase-shifts (outer).

While chords may contain combinations of four pitch classes or even more, they are frequently reduced to triads (i.e. sets of three pitch classes), with a predominance of major and minor triads. A minimal representation of triads boils down to 24 distinct triads (12 major and 12 minor). In the rest of this article, we restrict ourselves to this case, in spite of its simplified nature.

In order to model relations between triads, we consider triadic circles, i.e. circular arrangements of chords aimed at reflecting some proximity relationship between triads along their circumference.

The circle of thirds is formed by alternating major and minor triads with neighbouring triads sharing two common pitch classes, which is a way to model some kind of proximity between chords. In particular, chords belonging to a given key lie in a same sector of the circle of thirds. As an alternative, we also consider the circle of phase-shifts, which consists of a chord progression resulting from a minimal displacement on the 3-5 phase torus of triads as defined in [1]. Both circles are shown together on Fig. 2.

Each circle provides a way to express (in a unique way), the relationship between two triads, as the angular displacement along the circle. Note that a "chromatic" circle (... B_m B C_m C D^b_m D^b ...) could also be considered, but it is not represented on Fig. 2, for reasons explained later.

2.3 Systemic Organization

Based on the hypothesis that the relations between musical elements form a system of projective implications, the System & Contrast (S&C) model [2] has been recently formalized [3] as a generalization and an extension of Narmour's Implication-Realization model [16]. Its applicability to various music genres for multidimensional and multiscale music analysis has been explored in [7] and algorithmically implemented in an early version as "Minimal Transport Graphs" [10].

The S&C model primarily assumes that relations between 4 elements in a musical segment x_0 x_1 x_2 x_3 can be viewed as based on a two-scale *system* of relations rooted on the first element x_0 (the *primer*), which thus plays the role of a common *antecedent* to all other elements in the

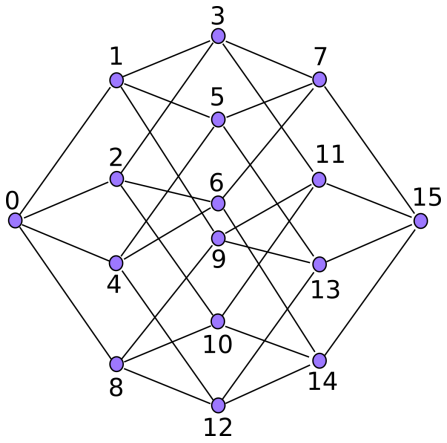

Figure 3. Tesseract where vertices at a same depth (or geodesic distance) from vertex #0 are aligned vertically. The resulting partial order between vertices is causal.

system. This is the basic principle that enables the joint modeling of several timescales simultaneously.

In the S&C approach, it is further assumed that latent systemic relations $x_1 = f(x_0)$ and $x_2 = g(x_0)$ trigger a process of projective implication:

$$x_0\ f(x_0)\ g(x_0) \overset{implies}{\Longrightarrow} g(f(x_0)) = \hat{x}_3 \qquad (1)$$

The *virtual* element $\hat{x}_3$ may then be more or less strongly denied by a *contrast*: $x_3 = \gamma(\hat{x}_3) \neq \hat{x}_3$, which creates some sense of closure to the segment.

In this work, the S&C model is used as the basic scheme to describe systems of music elements forming the faces of the tesseract.

3. GRAPH-BASED DESCRIPTION

3.1 Nested Systems

Elementary systems of four elements, as introduced in Section 2.3, can be used to describe longer sequences of musical events. In particular, sequences of 2^n elements arranged on an n-cube, provide a layout of the data where each face potentially forms a S&C, involving time instants that share specific relationships in the metrical grid.

As opposed to the sequential viewpoint which assumes a total order of elements along the timeline, the systemic organization on the tesseract leads to a partial order (illustrated on Fig. 3), where elements of the same depth are aligned vertically and where, in the framework of the S&C, the fourth element of each face can be defined in reference to the virtual element resulting from the projective implication of the three others. In the most general case, valid systemic organizations can be characterized by a graph of nested systems, the flow of which respects the partial ordering of Fig. 3. Note however that there is a possible conflict between three implications systems for elements 7, 11, 13 and 14 (each possible implication corresponding to a face of the tesseract [2]), and six for element 15.

[2] for instance, node 7 can be viewed as resulting from 3 implication systems: $[1, 3, 5, 7]$, $[2, 3, 6, 7]$ and $[4, 5, 6, 7]$.

$P0$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	A	A	A	A	B	B	B	B	C	C	C	C	D	D	D	D
$P1$	0	1	4	5	2	3	6	7	8	9	12	13	10	11	14	15
	A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D
$P2$	0	2	4	6	1	3	5	7	8	10	12	14	9	11	13	15
	A	B	A	B	A	B	A	B	C	D	C	D	C	D	C	D
$P3$	0	1	8	9	2	3	10	11	4	5	12	13	6	7	14	15
	A	A	B	B	C	C	D	D	A	A	B	B	C	C	D	D
$P4$	0	2	8	10	1	3	9	11	4	6	12	14	5	7	13	15
	A	B	A	B	C	D	C	D	A	B	A	B	C	D	C	D
$P5$	0	4	8	12	1	5	9	13	2	6	10	14	3	7	11	15
	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D

Table 1. List of the 6 PPPs, together with their corresponding Prototypical Carrier Sequences (PCS).

3.2 Primer Preserving Permutations (PPP)

One way to handle these conflicts is to constrain the graph to preserve systemic properties at higher scales. This can be achieved by forcing lower-scale systems to be supported by parallel faces on the tesseract, while the first elements of each of the 4 lower-scale systems are used to form an upper-scale system. This approach drastically brings down the number of possible graphs to 6, which corresponds to specific permutations of the initial index sequence (see Table 1), termed here as PPP (Primer Preserving Permutations).

To illustrate a PPP, let's consider the subdivision of a sequence of 16 chords into four sub-sequences of four successive chords. Each sub-sequence can be described as a separate Lower-Scale S&C (LS): $[0, 1, 2, 3]$, $[4, 5, 6, 7]$, $[8, 9, 10, 11]$ and $[12, 13, 14, 15]$. Then, these four S&Cs can be related to one another by forming the Upper-Scale S&C (US) $[0, 4, 8, 12]$, linking the four primers of the 4 LS. This configuration ($P0$) turns out to be particularly economical for describing chord sequences such as SEQ_1:

$Cm\ Cm\ Cm\ Bb \quad Ab\ Ab\ Ab\ Gm \quad F\ F\ F\ F \quad Cm\ Cm\ Bb\ Bb$

as most similarities develop between neighbouring elements.

If we now consider the following example (SEQ_2):

$Bm\ Bm\ A\ A \quad G\ Em\ Bm\ Bm \quad Bm\ Bm\ A\ A \quad G\ Em\ Bm\ Bm$

a different configuration appears to be more efficient to explain this sequence. In fact, grouping chords into the following 4 LS: $[0, 1, 8, 9]$, $[2, 3, 10, 11]$, $[4, 5, 12, 13]$ and $[6, 7, 14, 15]$, and then relating these four faces of the tesseract by a US $[0, 2, 4, 6]$ (configuration $P3$) leads to a less complex (and therefore more economical) description of the relations between the data within the segment. Fig. 4 illustrates these two configurations.

3.3 Prototypical Carrier Sequences (PCS)

Each of the 6 PPPs can be related to a prototypical structural pattern which turns out to be the one that is the most concisely described in the framework of this particular configuration. These 6 patterns, identified in Table 1, can be interpreted as "Prototypical Carrier Sequences" (PCS) over which the actual chord sequence appears as partially "modulated" information.

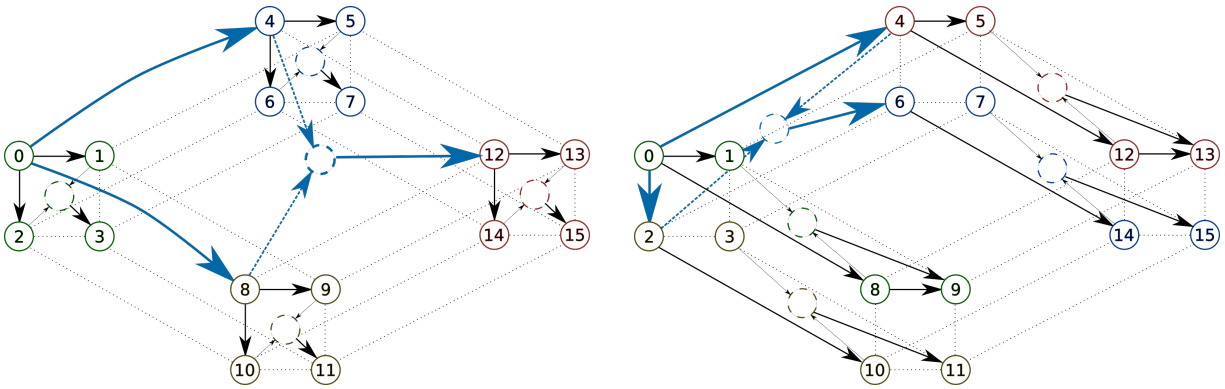

Figure 4. Representations of two PPP-based PGLRs on a tesseract: $P0$ (left), $P3$ (right). In blue, the Upper-Scale S&C – in black, the 4 Lower-Scale (LS) S&Cs. Dotted nodes indicate the virtual elements ($\hat{x}$) in the implication scheme (Section 2.3).

For instance, SEQ_1 appears merely as a sequence of type $P0$, which has been altered in positions 3, 7, 14 and 15 from the following carrier system:

$Cm\ Cm\ Cm\ Cm \quad Ab\ Ab\ Ab\ Ab \quad F\ F\ F\ F \quad Cm\ Cm\ Cm\ Cm$

Conversely, SEQ_2 exhibits a pattern that strongly relates to $P3$, with scattered deviations from:

$Bm\ Bm\ A\ A \quad G\ G\ Bm\ Bm \quad Bm\ Bm\ A\ A \quad G\ G\ Bm\ Bm$

located in positions 5 and 13.

Inferring the PCS shows interesting analogies with a demodulation operation and/or an error correcting code process, by concentrating the redundancy on the carrier sequence. It can of course happen that a sequence has several possible descriptions of equivalent plausibility, i.e. multiple coexisting interpretations w.r.t. its prototypical PPP structure.

In summary, PPP provide a limited set of baseline multiscale structural patterns which can be used to characterize actual chord sequences, via a minimum deviation criterion.

3.4 Algorithmical Considerations

In practice, given a (chord) sequence, $X = x_0 \dots x_{l-1}$, its optimal description ($\mathcal{D}^X$) within a class of PGLRs, can be obtained by minimizing a criterion $\mathcal{F}$ written as:

$$\mathcal{D}^X = [\Psi^X, R^X] = argmin_{[\Psi,R]}\, \mathcal{F}(\Psi, R|X) \quad (2)$$

where Ψ is a PGLR and R is a set of latent relations compatible with the connections of Ψ.

In the general case, both quantities are optimized jointly, considering all possible relations between each set of elements associated to each possible Ψ, and minimizing the cost over the whole sequence X.

Assuming that $\mathcal{F}$ is measuring the complexity of the sequence structure, $\mathcal{D}^X$ can be defined as the shortest description of the sequence. Therefore, searching for $\mathcal{D}^X$ can be seen as a Minimum Description Length (MDL) problem [21] and $\mathcal{F}$ can be understood as a function that evaluates the size of the "shortest" program needed to reconstruct the data [9]. This is strongly related to the concept of Kolmogorov complexity, which has received increasing interest in the music computing community over the past years [12, 14, 15, 19].

In the general case, the above optimization problem may turn out to be of a relatively high combinatorial complexity (see [10, 11]). But when considering triads over a circular arrangement, and limiting the set of possible Ψ to 6 PPPs, the optimization of $\mathcal{D}$ becomes easily tractable: all six PPPs can be tested exhaustively and for each of them, the set R comprises 16 relations (15 displacements over the triadic circle + the initialization of x_0) which are uniquely defined. Therefore, each cost can be readily computed as the sum of 16 terms, and the minimal PPP is easily found.

4. EXPERIMENTS

In order to study the ability of the PGLR model to capture structural information in chord sequences, we have carried out a set of experiments on the RWC POP dataset [8] on a corpus of 727×16 beat-synchronous chords sequences annotated manually as triads [3].

As there exists no ground truth as of the actual structure of a chord sequence, we compare different models as regards their ability to predict and compress chord sequences: in other words, how much side information is brought by the structure model, that saves information needed to describe of the content.

4.1 Distribution of PPPs

For each chord sequence X, the polytopic S&C graph P_X, corresponding to the PPP with minimal cost can be estimated by the optimization algorithm of Section 3.4. This yields the histogram depicted on Figure 5 [4].

Permutation $P3$ appears as the dominant one ($\approx 33\%$) and this may be related to the fact that its prototypical carrier sequence corresponds to a rather common "antecedent-consequent" form in music (especially, in pop music). Conversely, the least frequent PPP ($P2$), displays a frequency of occurrence below 5%. Somewhere in between, the 4 other permutations see their frequencies ranging loosely within 10% to 20%).

[3] Data are available on *musicdata.gforge.inria.fr/RWCAnnotations.html*

[4] About $2/3$ of test sequences correspond to a unique optimal PPP but when $k > 1$ permutations provide equally optimal solutions, each of them is counted as $1/k$.

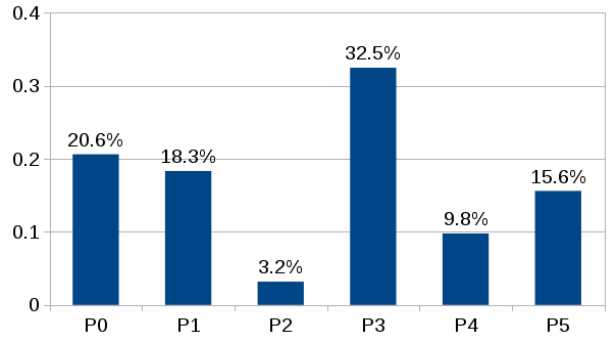

Figure 5. Histogram of PPPs across the test data.

4.2 Prediction and Compression

In order to compare the prediction and compression capabilities provided by multiscale polytopic graphs, we consider 4 structure models:

- S, a purely sequential graph where each element is related to its immediate predecessor [5],
- P_0, the polytopic S&C graph corresponding to PPP P_0 for all sequences,
- P_3, the polytopic S&C graph corresponding to PPP P_3 for all sequences,
- P_X, the polytopic S&C graph corresponding to the PPP with minimal cost optimised a posteriori for each 16-chord sequence, as described in section 3.4.

All models are first-order models, in the sense that any given element within the sequence is related to a single antecedent (its time predecessor for the sequential graph, a primer or a virtual element, in the case of the S&C model).

Performance for each model is obtained by calculating a *perplexity* [4] B^*, derived from the *entropy* H^*.

Given a model M, the computation of perplexity requires the definition of a probability density function (pdf) for all observable events which underlie the model. In our case, this means assigning a probability value $P_M(x_i|\Phi(x_i))$ to any pair $(x_i, \Phi(x_i))$, where $\Phi(x_i)$ is the antecedent of x_i in the graph. This is equivalent to defining $P_M(r(\Phi(x_i), x_i))$, where $r(\Phi(x_i), x_i)$ is the relation which turns $\Phi(x_i)$ into x_i. Considering a simple rotation angle $\theta(x_2|x_1) = \theta_2 - \theta_1$ on the triadic circle, $P_M(r(x_1, x_2))$ is a pdf that takes $z = 24$ distinct values.

The entropy of model M can be computed as:

$$H^*(M) = -\sum_{k=0}^{z} P_M(r_k) \log_2 P_M(r_k) \qquad (3)$$

$B^* = 2^{H^*}$ can be interpreted as a branching factor, that is the equivalent number of distinct relations between two chords, if these relations were equiprobable. It measures the compression capacity of the model and is smaller for models which capture more information in the data.

[5] This corresponds to a sequential bi-gram model, a very common approach in MIR [18].

	Triad Circle		
	Third	Phase	Random
$B(S)$	8.00	7.67	9.32
$B(P_0)$	6.68	6.77	7.84
$B(P_3)$	5.35	5.35	6.02
$B(P_X)$	**4.63**	**4.63**	**5.21**
$B_{tot}(P_X)$	5.18	5.18	5.83

Table 2. Average cross-perplexity obtained for the various models on RWC-Pop data with 2-fold cross-validation (training on even songs + testing on odd songs and vice-versa).

In this work, we consider specifically the *cross-perplexity* B derived from the *negative log likelihood* (NLL) $\hat{H}$, computed on a test-set (of L observations). In that case, the capacity of the model to catch relevant information from an unseen musical segment is measured by means of a cross-entropy score, which quantifies the ability of the model to predict unknown sequences from a similar (yet different) population.

For a given model M, $\hat{H}$ is defined as:

$$\hat{H}(M) = -\frac{1}{L}\sum_{i=0}^{L-1} \log_2 P_M(x_i|\Phi(x_i)) \qquad (4)$$

with the convention $P(x_0|\Phi(x_0)) = 1/24$.

In that context, the cross-perplexity $B = 2^H$ can be understood as an estimation of the (per symbol) average branching factor in predicting the sequence knowing its structure, on the basis of probabilities learnt on *other* sequences, assumed to be of the same sort.

Additionnally, for model P_X, we also compute the total entropy $\hat{H}_{tot}(P_X) = \hat{H}(P_X) + Q$, which includes the number of bits needed to encode the optimal configuration of the PPP (1 among 6) for each sequence of 16 chords, namely:

$$Q = \log_2(6)/16 \approx 0.16 \text{ bits/symbol}, \qquad (5)$$

this term being equal to 0 for the other models.

The first column in Table 2 compares cross-perplexity figures obtained with the 4 structure models and considering the circle of thirds for modeling relations between chords. These results show that all tested polytopic models outperform the sequential model, with an additional advantage for the P_X approach, even when taking into account the overhead cost required for PPP configuration encoding.

4.3 Impact of the Triadic Circle

In the rest of Table 2, cross-perplexity values are provided for two other circles of triads: the circle of phase-shifts as defined on Fig. 2 and a randomized circle, where triads are positioned at random. Results show that the phase circle performs quite the same as the circle of thirds, whereas the randomized circle clearly performs less well. All outperform their counterpart in the sequential model, as for all polytopic models, the identity relation is of zero cost and higher probability. We do not report perplexities on the chromatic circle, given that it is congruent to the circle of thirds, thus yielding strictly identical results.

US	LS_1	LS_2	LS_3	LS_4
44.4 %	8.0 %	15.7 %	19.7 %	22.4 %

Table 3. Proportion of sequences with contrastive US (Upper-Scale system) and LS_k (k^{th} Lower-Scale system).

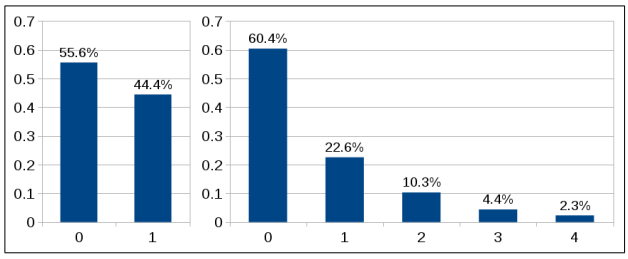

Figure 6. Proportion of contrastive systems within US systems (left) and the 4 LS systems (right)

4.4 Distribution and Density of Contrasts

To study the specific relationship between the virtual and the contrastive element in the P_X scheme, we investigated on the location and the number of contrastive vs. non-contrastive elements in potentially contrastive positions defined by the PPP framework.

Table 3 presents the distribution of actual contrasts for the Upper-Scale (US) and the 4 Lower-Scale (LS) in contrastive positions. While 44.4% of Upper-Scale Systems are contrastive, it can also be noted that the frequency of LS contrasts (or, so to speak, the occurrence of surprises at the lower-scale span) increases with the index of the LS system (i.e., its depth in the tesseract).

Figure 6 depicts the proportion of sequences as a function of the number of actual contrasts observed in different contrastive positions. It can be observed that the number of contrastive Lower-Scale systems decays (roughly exponentially) from over 60.4% of sequences with no contrastive Lower-Scale system down to only 2.3% with all 4 LS systems being contrastive.

It would surely be interesting to compare these profiles across different music genres and a variety of musical dimensions, in order to study possible correlations.

4.5 Contrast Intensity

Table 4 reports the perplexity obtained when considering separately the systemic positions and the contrastive positions. Keeping in mind that they may be specific to the corpus, results show nevertheless two very interesting trends.

Perplexity is higher in systemic positions (5.6) as opposed to constrastive positions (3.5), implying that the actual observations in contrastive positions often correspond (or are close) to the projective implication. This can be related to the results observed in the previous section, w.r.t. the relatively low density of actual contrasts.

However, when different from identity (column Diff), these relations show a *lower* perplexity for systemic relations (14.6 vs 18.7) indicating that, when a relation is not identity, the contrast is more unpredictable and/or more distant on the circle of thirds, than it is for systemic relations.

	All	Diff
Systemic position	5.6	14.6
Contrastive position	3.5	18.7

Table 4. Perplexity of relations for systemic relations and contrastive relations, including (All) or excluding (Diff) the identity relation.

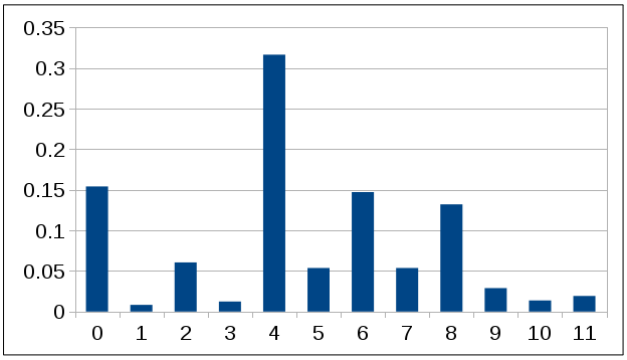

Figure 7. Proportion of chord sequences showing n distortions relative to their Prototypical Carrier Sequence (PCS).

In summary, strictly contrastive relations tend to be less frequent but more intense than systemic relations. This certainly relates to the presumed role of contrasts as carrying a strong quantity of surprise. These observations may be a motivation for a different treatment of systemic relations vs. contrastive ones.

4.6 Distortion of Prototypical Carrier Sequences

Ultimately, we considered the distribution of the number of distortions between observed chord sequences and their PCS, as defined in section 3.3. Figure 7 shows a dominance of 4 deviations, with an overall prevalence of even values, suggesting that modelling systems of *relations* (i.e. edges) within the tesseract could be useful to further improve the compression capabilities of the PGLR model.

5. CONCLUSIONS

Both from the conceptual and experimental viewpoints, the polytopic approach presented in this article appears as an efficient way to model multiscale relations in chord sequences.

While still at an early stage of development, the PGLR model provides a potentially useful and powerful framework for a number of tasks in MIR, as well as interesting tracks for music analysis and generation. Indeed, the core principles of the PGLR scheme are not specific to chord sequences: its application to other types of musical objects, such as melodic motives and rhythmic patterns are currently being explored.

Ongoing work also includes the extension of the polytopic model to a wider range of timescales, and the handling of segments of irregular size.

6. REFERENCES

[1] Emmanuel Amiot. *The Torii of Phases*, pages 1–18. Springer, Mathematics and Computation in Music: 4th International Conference, MCM 2013, Montreal, QC, Canada, June 12-14, 2013. Proceedings, Berlin, Heidelberg, 2013.

[2] Frédéric Bimbot, Emmanuel Deruty, Gabriel Sargent, and Emmanuel Vincent. Semiotic structure labeling of music pieces: concepts, methods and annotation conventions. In *Proc. ISMIR*, 2012.

[3] Frédéric Bimbot, Emmanuel Deruty, Gabriel Sargent, and Emmanuel Vincent. System & Contrast : A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces. *Music Perception*, 33:631–661, June 2016. Former version published in 2012 as Research Report IRISA PI-1999, hal-01188244.

[4] Peter F Brown, Vincent J Della Pietra, Robert L Mercer, Stephen A Della Pietra, and Jennifer C Lai. An Estimate of an Upper Bound for the Entropy of English. *Computational Linguistics*, 18(1):31–40, 1992.

[5] Clifton Callender, Ian Quinn, and Dmitri Tymoczko. Generalized voice-leading spaces. *Science*, 320(5874):346–348, 2008.

[6] Richard Cohn. *Audacious Euphony: Chromatic Harmony and the Triad's Second Nature*. Oxford University Press, 2011.

[7] Emmanuel Deruty, Frédéric Bimbot, and Brigitte Van Wymeersch. Methodological and Musicological Investigation of the System & Contrast Model for Musical Form Description. Research Report RR-8510, INRIA, 2013. hal-00965914.

[8] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka. RWC Music Database: Popular, Classical and Jazz Music Databases. In *ISMIR*, volume 2, pages 287–288, 2002.

[9] Peter Grunwald and Paul Vitányi. Shannon Information and Kolmogorov Complexity. *arXiv preprint cs/0410002*, 2004.

[10] Corentin Louboutin and Frédéric Bimbot. Description of Chord Progressions by Minimal Transport Graphs Using the System & Contrast Model. In *ICMC 2016 - 42nd International Computer Music Conference*, Utrecht, Netherlands, September 2016.

[11] Corentin Louboutin and Frédéric Bimbot. Polytopic Graph of Latent Relations: A Multiscale Structure Model for Music Segments. In *MCM 2017 - 6th International Conference of the Society of Mathematics and Computation in Music*, Mexico City, Mexico, June 2017.

[12] Corentin Louboutin and David Meredith. Using General-Purpose Compression Algorithms for Music Analysis. *Journal of New Music Research*, 45(1):1–16, 2016.

[13] Matthias Mauch, Katy Noland, and Simon Dixon. Using Musical Structure to Enhance Automatic Chord Transcription. In *ISMIR*, pages 231–236, 2009.

[14] Panayotis Mavromatis. Minimum Description Length Modelling of Musical Structure. *Journal of Mathematics and Music*, 3(3):117–136, 2009.

[15] David Meredith. Music analysis and Kolmogorov complexity. *Proceedings of the 19th Colloquio d'Informatica Musicale (XIX CIM)*, 2012.

[16] Eugene Narmour. *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. University of Chicago Press, 1992.

[17] Hélène Papadopoulos and George Tzanetakis. Modeling chord and key structure with markov logic. In *ISMIR*, pages 127–132, 2012.

[18] Marcus Thomas Pearce. *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. City University London, 2005.

[19] David Temperley. Probabilistic Models of Melodic Interval. *Music Perception*, 32(1):85–99, 2014.

[20] Dmitri Tymoczko. Scale Theory, Serial Theory and Voice Leading. *Music Analysis*, 27(1):1–49, 2008.

[21] Paul MB Vitányi and Ming Li. Minimum description length induction, Bayesianism, and Kolmogorov complexity. *IEEE Trans. Information Theory*, 46(2):446–464, 2000.

[22] Xinquan Zhou and Alexander Lerch. Chord Detection Using Deep Learning. In *Proceedings of the 16th ISMIR Conference*, volume 53, 2015.