

Table des figures.....	vi
Introduction	1
Objet du travail.....	1
Périmètre de l'étude.....	2
Modèles et hypothèses.....	2
Objectifs parallèles	3
Chapitre 1 Fondements et Positionnement	5
1.1 Hypothèse de travail par rapport à la structure musicale.....	5
1.1.1 Introduction	5
1.1.2 Domaine musicologique et technologique	6
1.1.3 Hypothèses de travail	7
1.1.4 Choix du style et éléments musicaux	9
1.1.5 Critère et méthodologie	10
1.2 Système d'implication et modèle Système & Contraste (S&C).....	11
1.2.1 Système d'implication	11
1.2.2 Système & Contraste	13
1.3 Généralisation multi-dimensionnelle et graphes polytopiques.....	15
1.4 Variations et Recompositions Polytopiques.....	19
1.4.1 Discussion par rapport aux variations existantes.....	19
1.4.2 Recompositions Polytopiques	20
1.5 Théorie de l'Information et complexité.....	23
1.6 Méthodologie.....	24
1.6.1 Procédé de recomposition.....	25
1.6.2 Algorithmes de segmentation	25
1.6.3 Caractérisation des variations polytopiques	25
1.6.4 Validation empirique	26
1.6.5 Conception d'un outil créatif.....	26
Chapitre 2 Concepts et Formalisme.....	27
2.1 Objets musicaux & échelles temporelles.....	27
2.1.1 Forme réduite de l'information musicale	31
2.1.2 Surface musicale et segmentation.....	33
2.1.3 Graphe tesseractique.....	38
2.1.4 Systèmes d'implications au sein du modèle polytopique.....	39
2.2 Recompositions Polytopiques	44
2.2.1 Définition.....	44
2.2.2 Les PPP : des recompositions particulières	45
2.2.3 Caractérisation des recompositions	48
2.3 Conclusion.....	49
Chapitre 3 Critère de complexité et algorithme de segmentation	51
3.1 Critère de complexité.....	52
3.1.1 Complexité d'une section musicale.....	53

3.1.2 Estimation de la complexité entre les systèmes.....	54
3.1.3 Complexité d'un système	56
3.1.4 Complexité conditionnelle d'un motif.....	58
3.1.5 Distance entre deux granules.....	65
3.2 Inférence d'une segmentation optimale pour une méthode d'estimation donnée.....	67
3.2.1 Caractérisation d'une méthode algorithmique.....	67
3.2.2 Détermination de θ^* et du score de complexité.....	69
3.2.3 Calcul du score de complexité d'une segmentation	71
3.2.4 Combinatoire, temps de calcul et optimisation	72
3.2.5 Détails d'implémentations.....	73
3.3 Méthodes algorithmiques et évaluation préliminaire	75
3.3.1 Génération des recompositions polytopiques.....	75
3.3.2 Annotations : segmentations manuelles par deux personnes du domaine	76
3.3.3 Evaluation par rapport aux annotations manuelles.....	77
3.3.4 Résultats préliminaires	80
Chapitre 4 Tests perceptifs	85
4.1 Test 1 : évaluation des recompositions polytopiques vs aléatoires	86
4.1.1 Données	86
4.1.2 Protocole.....	88
4.1.3 Évaluation des scores.....	90
4.1.4 Résultats	91
4.2 Test 2 : évaluation perceptive des segmentations algorithmiques.....	94
4.2.1 Données	94
4.2.2 Protocole de sélection des variantes de méthodes.....	95
4.2.3 Sélections des méthodes algorithmiques testées	96
4.2.4 Protocole du test perceptif.....	98
4.2.5 Évaluation des scores.....	98
4.2.6 Résultats	99
4.3 Conclusion test perceptif	104
Chapitre 5 Interface de Création Musicale Par Recomposition Polytopique	105
5.1 Interfaces de variations existantes	105
5.2 Application Tessy	105
5.2.1 Modes de fonctionnement	106
5.2.2 Variations polytopiques.....	108
5.2.3 Visualisation temporelle.....	109
5.3 Perspectives de l'application Tessy.....	111
Chapitre 6 Conclusions et débouchés.....	113
Bibliographie	115

Table des figures

Figure 1: Forme carrée du modèle Système & Contraste	13
Figure 2 : Systèmes carrés d'implication	13
Figure 3: Généralisation multi-échelle des systèmes d'implications	15
Figure 4: Présentations des relations structurelles entre éléments à différentes échelles temporelles.....	16
Figure 5: Passage de la section musicale en vue séquentielle à une vue polytopique de dimension 4.....	17
Figure 6: Exemples de parcours du polytope support hypercube dans un ordre non séquentiel.	18
Figure 7 : Mélodie du refrain de Adele – Skyfall.	20
Figure 8 : Recomposition aléatoire de la mélodie du refrain de Adele - Skyfall.....	20
Figure 9 : Illustration des notes en position métrique homologues à la note n°1 à différentes échelles métriques (1/2 mesure, 1 mesure, 2 mesures et 4 mesures).	21
Figure 10: Représentation des 6 systèmes de face parallèles sur le tesseract et leur permutation PPP correspondante.	22
Figure 11: Tables des niveaux d'expériences musicales extrait de (Snyder, 2000).	28
Figure 12: Table récapitulative des échelles manipulées dans ces travaux.	30
Figure 13: Vue Piano Roll de la mélodie originale.....	32
Figure 14: Vue Piano Roll de la mélodie avec ses notes prolongées.....	32
Figure 15 : Vue Piano Roll de la mélodie réduite après échantillonnage en granules.....	32
Figure 16: Portée d'une section musicale avec une surface de 8 mesures en version réduite.	33
Figure 17: Illustration du phénomène d'anacrouse et de débordement sur la partition de « Joyeux Anniversaire ». Les barres diagonales, décalées par rapport aux barres de mesure, figurent les regroupements de notes en motifs au niveau de la surface musicale. ..	34
Figure 18: Illustration des éléments granulaires constituant les motifs structurants. Ici, les 6 premières croches sont laissées hors du périmètre de modélisation du segment.	35
Figure 19: Segmentation régulière anticipée de 1/4 mesure ($\theta = -2$ granules = -1 temps) ..	37
Figure 20. Segmentation régulière débutant sur la première mesure ($\theta = 0$ granule = 0 temps)	37
Figure 21: Segmentation régulière retardée de 1/2 mesure ($\theta = 4$ granules = 2 temps).....	37
Figure 22: Segmentation X0 en 16 motifs de la section X.	38
Figure 23: Visualisation des 16 motifs de la section sur son graphe polytopique (ici un tesseract)	38
Figure 24: Illustration des relations présentes au sein d'un système.....	39
Figure 25: Représentation polytopique des 4 systèmes d'implications formés par les motifs dans l'ordre séquentiel.	40
Figure 26: Représentation polytopique des 4 systèmes d'implications formés par les motifs selon un des systèmes de faces parallèles.....	41
Figure 27 : Choix des directions permettant de caractériser les niveaux des systèmes d'implication dans le tesseract.....	42
Figure 28: Représentation des 6 systèmes de faces parallèles du tesseract.	42
Figure 29: Illustration des Super Systèmes pour chacun des systèmes de faces parallèles.	43
Figure 30: Définition des permutations (polytopiques) préservant les primers (PPP).....	45
Figure 31: Illustration des parcours du tesseract selon les différentes PPP.	46

Figure 32: Exemple des 5 recompositions polytopiques PPP appliquées à la mélodie d'un extrait musical de la base RWC POP.....	47
Figure 33: Illustration des 7 discontinuités induites par la recomposition π_2	48
Figure 34: Illustration de la caractéristique d'excursion maximale $E = 3$, induite par la fonction de recomposition π_2	49
Figure 35: Caractéristiques topologiques des 6 PPP.....	49
Figure 37: Illustration du calcul de complexité lorsque les 4 systèmes d'implications sont considérés indépendants.....	54
Figure 38: Illustration du calcul de complexité lorsque l'on prend en compte la complexité liée au Super-Système.....	55
Figure 39: Relations de dépendances au sein d'un système d'implication pour l'estimation de sa complexité.	56
Figure 40: Illustration de l'estimation d'un système d'implication par rapport au premier élément.	57
Figure 41: Illustration de l'estimation d'un système d'implication avec recherche du meilleur antécédent de l'élément contraste.....	57
Figure 42: Représentation des motifs x et y constitué respectivement de m et n granules.....	58
Figure 43: Exemple de deux motifs x et y dans leur représentation Piano Roll.	59
Figure 44: Expression de la complexité C_{yx} sous l'hypothèse de l'indépendance des granules constituants.....	59
Figure 45: Exemple de deux motifs identiques à un décalage fréquentiel près.....	60
Figure 46: Représentation de l'estimation de la complexité $C(y x)$ selon l'hypothèse de propagation du premier intervalle δ_1	61
Figure 47: Expression de la complexité $C(y x)$ selon l'hypothèse de propagation d'intervalle de chaque couple de granules.....	62
Figure 48: Deux motifs où le temps fort coïncide avec le début du granule 3.	63
Figure 49: Illustration du principe d'estimation bilatérale avec propagation cascade	64
Figure 50: Distance entre deux granules.....	65
Figure 51: Tableau récapitulatif des variantes de distances entre granules considérées dans cette thèse. d_c : distance chromatique – d_m : distance musicale	66
Figure 52: Table des caractéristiques des méthodes d'estimations, définissant l'ensemble M des méthodes Γ	68
Figure 53: Exemple de segmentation selon la grille régulière et un décalage θ de -2 temps.....	70
Figure 54: Schéma explicatif du calcul de complexité avec dépendances entre objets musicaux à 4 échelles temporelles.....	72
Figure 55: Exemple d'affichage des notes d'un fichier MIDI avec music21.	74
Figure 56: Interface de segmentation manuelle pour la collecte des annotations de référence. Les boutons colorés permettent d'écouter les recompositions polytopiques selon la grille choisie.	77
Figure 57: Exemple de 2 méthodes appariées avec une différence sur leur 1 ^{ère} caractéristique.	79
Figure 58: Table des résultats de l'évaluation de concordance des caractéristiques algorithmiques	81
Figure 59: Configuration restantes après élimination de la caractéristique « distance chromatique » entre granules.....	82
Figure 60 : Concordance des méthodes algorithmiques avec le jeu d'annotation A1	82
Figure 61 : Concordance des méthodes algorithmiques avec le jeu d'annotation A2	82
Figure 62: Exemples de partitions de recompositions RGP et PPP	87

Figure 63: Instructions données aux sujets participant aux tests perceptifs.	89
Figure 64: Echelle de notation des recompositions dans le cadre des tests perceptifs.	89
Figure 65: Résultats des tests perceptifs par sujet sur la question de la cohérence musicale.	92
Figure 66: Résultats des tests perceptifs par sujet sur la question de l'esthétique musicale.	92
Figure 67: Comparaison des résultats des tests perceptifs en fonction du type de recomposition (critère de cohérence musicale).....	93
Figure 68: Comparaison des résultats des tests perceptifs en fonction du type de recomposition (critère de l'attractivité musicale).	93
Figure 69 : Récapitulatif des méthodes évaluées dans le second test perceptif.	97
Figure 70: Tableau présentant les méthodes algorithmiques et leurs scores de cohérence (Second test perceptif)	101
Figure 71: Résultats des scores de cohérence du second test perceptif	101
Figure 72: Tableau présentant les méthodes algorithmiques et leurs scores d'attrait (Second test perceptif)	103
Figure 73: Résultats des scores d'attractivité du second test perceptif.....	103
Figure 74: Présentation de l'interface de l'application Tessy	106
Figure 75: Illustration de l'application Tessy en mode lecture par segment.	107
Figure 76: Présentation de la logique de "complétion automatique" pour conservation de la logique systémique dans Tessy	107
Figure 77: Interface de contrôle permettant de lancer une séquence PPP dans l'application Tessy.	108
Figure 78: Présentation des cases mémoires dans l'application Tessy.	109
Figure 79: Présentation des deux modes de visualisation temporelle de l'application Tessy sur la même recomposition PPP.	109
Figure 80: Réglages avancés de l'application Tessy	110

Introduction

Objet du travail

Le travail présenté dans ce manuscrit s'intéresse à l'exploration, la validation et l'utilisation d'un modèle de structure musicale à des fins scientifiques et créatives.

La définition de la structure musicale est un sujet largement étudié et débattu, autant dans la communauté musicologique que dans le domaine de la recherche d'information musicale (MIR). Cependant, il n'existe toujours pas, à ce jour, une conception unique et universellement admise de ce qui définit la structure musicale.

Dans ce travail, on a choisi une approche expliquant la narration d'une phrase musicale comme une organisation entre éléments similaires et dissimilaires qui constitue un flux d'information fait de redondances et de surprises.

La musique est habituellement considérée comme un phénomène séquentiel, où les éléments, les motifs et autres groupements de notes se produisent selon une dimension chronologique. La musique est ainsi principalement vue comme un flux d'information dont l'organisation temporelle serait essentiellement due aux relations entre éléments adjacents.

Cependant, on observe que les éléments musicaux présentent également des relations de similarité ou de dissimilarité avec d'autres éléments situés à distance dans la narration musicale et à plusieurs échelles de temps différentes. Toutefois cet aspect multi-échelle est rarement pris en compte dans les modèles de flux d'information musicale qui opèrent principalement de proche en proche.

Avant de continuer, on notera que la nature multi-échelle de la musique est à distinguer de sa nature multidimensionnelle, reposant sur l'existence et les rapports entre de multiples dimensions musicales telles que le rythme, l'harmonie, la mélodie, etc. Ceci n'est pas l'objet du présent travail.

Périmètre de l'étude

Dans les travaux de cette thèse, on s'intéresse tout particulièrement à la structure mélodique. Pour une importante proportion de morceaux de musique, la mélodie est un élément structurant essentiel et il est donc particulièrement intéressant d'étudier son organisation et sa construction. On remarque également que la mélodie est particulièrement adaptée à la problématique étudiée dans cette thèse du fait des nombreuses redondances qu'elle fait apparaître au fil du temps.

Plus particulièrement, notre sujet d'étude se concentre sur la musique dite « Pop ». Une grande majorité de la musique produite et écoutée aujourd'hui appartient à ce style musical. De plus sa construction a tendance à être relativement codifiée par rapport à d'autres genres musicaux et constitue donc sujet d'étude privilégié pour tester et évaluer des modèles simples de la structure musicale avant d'essayer de les éprouver sur des styles musicaux plus vastes. Enfin, jusqu'à présent, l'essentiel des études sur la structure musicale se sont intéressées à l'agencement entre sections : introduction, couplet, refrain, pont, etc. Dans le présent travail, nous nous concentrons sur l'organisation interne des sections et plus particulièrement sur la structure des refrains.

Modèles et hypothèses

Afin de représenter la structure multi-échelle d'une mélodie dans un refrain, on s'appuie sur une représentation dite à support polytopique, qui constitue un cadre d'analyse et de représentation de relations structurelles adapté aux données musicales.

Différents travaux, et notamment ceux de Louboutin (Louboutin & Bimbot, Polytopic Graph of Latent Relations: A Multiscale Structure Model for Music Segments, 2017), ont formalisé le modèle polytopique comme un graphe de relations latentes que nous utilisons dans ce travail pour décrire une section musicale. Dans ce graphe, chaque nœud correspond à un motif musical élémentaire de l'extrait et chaque arête porte la relation entre certains couples de motifs. Toutefois, en situation réelle, l'instanciation optimale des nœuds du graphe par les motifs musicaux considérés comme structurants est inconnue et elle constitue une étape essentielle et complexe.

Une hypothèse forte du modèle polytopique est de considérer qu'une part importante de la structure d'une phrase musicale est due aux redondances qui surviennent à des instants métriquement homologues, ce qui confère une pulsation d'information propre au contenu musical et détectable comme « porteuse structurante ».

En d'autres termes, on considère que c'est grâce à ce phénomène qu'un auditeur « lambda » perçoit la structure musicale d'un point de vue cognitif, même s'il n'est pas musicien, musicologue, ou expert du style musical dont est issu le passage musical.

Ainsi, une hypothèse du présent travail est l'existence d'un critère sous-jacent au caractère structuré d'un passage musical en lien avec la complexité de l'explication des relations entre les motifs qui le constitue. Cette idée revient à aller puiser dans les concepts de la théorie de

l'information et notamment la complexité de Kolmogorov, pour rendre compte de certains aspects de la structure musicale.

Plus concrètement, on vise à évaluer si la complexité du système de relation entre éléments musicaux assignés à un graphe polytopique est corrélée à la perception de sa structure.

L'approche poursuivie consiste à mettre en place différents scénarios de recomposition d'une section musicale d'origine et d'évaluer l'impact de ces recompositions sur leur pertinence et leur acceptabilité structurelle. Ces tâches de recomposition constituent un point central dans cette thèse.

Faute de disposer de critères objectifs permettant de quantifier la cohérence structurelle d'un passage musical, on met en place un protocole expérimental de tests perceptifs visant à obtenir des évaluations subjectives de ces recompositions auprès de panels d'auditeurs humains.

L'analyse des résultats de cette étude permet de dégager quelques tendances sur la pertinence et le potentiel du modèle de structure multi-échelle polytopique.

Objectifs parallèles

Complémentairement à cet objectif, il faut souligner que le principe de « variation » est depuis toujours au cœur de la création musicale, que ce soit les réarrangements par un même musicien des morceaux qu'il a composés ou pour apporter des variations d'œuvres pré-existantes. Ces variations consistent en général en un changement de certains paramètres musicaux comme la mélodie, le rythme, l'harmonie, l'instrumentation, voire l'agencement des sections musicales successives (comme dans le cas des « remix »), tout en respectant certaines contraintes pour préserver la musicalité du résultat.

Ici, on propose d'effectuer un nouveau type de variation qui consiste à créer un nouveau contenu en perturbant l'ordre temporel des motifs au sein d'une section, mais tout en préservant sa structure polytopique. Les recompositions obtenues conservent les systèmes structurants de l'extrait d'origine, mais en intervertissant les échelles hyper-métriques. On désigne ce processus par variation (et/ou recomposition) polytopique, lequel s'avère posséder un fort potentiel d'aide à la création musicale.

Enfin, une contribution de cette thèse est d'étudier et mettre à disposition une application permettant l'utilisation effective de ce paradigme à des fins créatives. Nous avons donc travaillé en partenariat avec l'entreprise Kercodex qui développe des applications mobiles, afin de réaliser une interface permettant l'utilisation du modèle pour générer des variations d'un contenu musical original. La mise à disposition de cette application (en tant que « preuve de concept ») est destinée à permettre au grand public de s'emparer des idées développées dans cette thèse, afin de composer et recomposer de la musique d'une façon nouvelle grâce aux représentations polytopiques, ce qui, à notre connaissance, n'avait jamais été exploré jusqu'à présent.

Chapitre 1

Fondements et Positionnement

1.1 Hypothèse de travail par rapport à la structure musicale

1.1.1 Introduction

D'une façon très générale, la structure musicale est définie comme l'organisation ou l'agencement d'une narration musicale. Le langage musical est en effet organisé selon une syntaxe et repose sur des éléments constituants qui sont agencés d'une manière plus ou moins structurée.

Il existe de nombreuses façons d'aborder la structure musicale et de multiples travaux traitent de ce sujet dans plusieurs branches de recherches, en musicologie par exemple ou en recherche d'information musicale.

La façon même d'appréhender la musique varie en fonction de nombreux facteurs. Tout d'abord le style musical, c'est-à-dire son époque, sa souche culturelle, son genre. Dans les travaux de cette thèse, on s'intéresse à la musique tonale occidentale actuelle et plus particulièrement à la musique dite 'Pop'.

Définir ce qu'est la structure musicale nécessite de préciser quels sont les aspects du processus musical que l'on considère comme prédominants à cet égard.

Comme évoqué dans (Côté-Lapointe, 2013) une première approche, dite « intrinsèque », considère les éléments musicaux comme phénomène en soi, sans tenir compte du contexte de création, d'écoute et de réception.

A l'inverse, l'approche « extrinsèque » intègre l'auditeur et son milieu comme facteurs importants dans la compréhension et description de la narration musicale. Cette approche inclut des considérations sociologique et psychologique, et fait intervenir notamment la prise en compte de conventions de composition musicale inhérentes aux contenus musicaux, pour un genre donné.

Dans les travaux de cette thèse, et dans la lignée directe de (Narmour, *The analysis and cognition of melodic complexity: The implication-realization model*, 1992), nous nous basons sur une conception plutôt de type intrinsèque, sans pour autant ignorer certaines dimensions liées à l'auditeur : nous relierons la structure musicale à l'existence de répétitions et/ou de similarités dans les éléments structurels du fait qu'elles provoquent des attentes cognitives de la part de l'auditeur dont la réalisation ou le déni constituent l'ossature narrative.

1.1.2 Domaine musicologique et technologique

Dans le domaine de la recherche musicologique, il existe un nombre important de travaux qui formalisent la structure musicale, par exemple les travaux de Schenker (Schenker, 1935). Dans le cadre Schenkerien, la structure de la musique tonale présente plusieurs niveaux : une structure profonde nommée l'Urzatz et une succession de formes de structures intermédiaires qui finalement aboutissent à la forme de la partition.

D'autres travaux, comme ceux de Lerdahl et la théorie GTTM (Generative Theory of Tonal Music) (Lerdahl, 1985) s'appuient sur un formalisme supposé répliquer des processus mentaux selon lesquels un auditeur construit sa compréhension de la musique selon 4 niveaux de structure. Les objets structurels de l'approche GTTM sont des arbres qui décrivent séquentiellement des dépendances hiérarchiques entre éléments musicaux.

Depuis plusieurs décennies, on a vu apparaître également beaucoup de travaux issus de la communauté en Recherche d'Information Musicale (MIR), comme les travaux de Peeters et al. sur la représentation de la structure par des matrices de similarités (Peeters, 2007), les travaux de Smith et al. pour l'annotation structurelle de grande base de données (Smith J. B., 2011), ou encore les travaux de Bimbot et al. sur la définition et les méthodologies d'annotation de la structure sémiotique (Bimbot, 2012).

On notera également la tenue de séminaires spécifiquement sur le sujet, par exemple le séminaire Dagstuhl de 2016 sur « Computational Music Structure Analysis » qui a rassemblé des chercheurs issus de différentes branches afin de discuter l'état de l'art du domaine et identifier et analyser les limites des méthodologies existantes pour le traitement informatique de la structure musicale.¹

Toutefois, il n'existe pas à ce jour de consensus scientifique ayant abouti à une définition communément admise de ce qu'est la structure musicale, qui reste donc une question ouverte et un domaine de recherche à bien des niveaux.

¹ *This triggered interdisciplinary discussions that leveraged insights from fields as disparate as psychology, music theory, composition, signal processing, machine learning, and information sciences to address the specific challenges of understanding structural information in music. Finally, we explored novel applications of these technologies in music and multimedia retrieval, content creation, musicology, education, and human-computer interaction*

1.1.3 Hypothèses de travail

Faute d'une conception commune, nous avons choisi, dans les travaux de cette thèse, une approche basée sur la théorie de l'information. Dans un cadre général, la Théorie de l'Information vise à quantifier le contenu en information d'un message ou d'un ensemble de messages. Les concepts de la Théorie de l'Information sont couramment utilisés en codage de l'information, en compression de données et en cryptographie par exemple.

L'étude de la structure musicale sous l'angle de la Théorie de l'Information est un axe de recherche de l'équipe Panama (précédemment Metiss) de l'Irisa de Rennes depuis une dizaine d'années et a donné lieu à une série de travaux explorant l'utilité de ces concepts dans différents contextes : annotation (Bimbot, 2012), segmentation structurelle de signaux musicaux (G. Sargent, 2017), modélisation de progressions harmoniques mélodiques ou rythmiques (Guichaoua, 2017), (Louboutin, 2019).

Dans le cadre de ces travaux, on considère que l'organisation d'une section de musique repose largement sur des alternances entre éléments présentant soit des similarités, soit des nouveautés, apportant tantôt de la redondance et tantôt de la surprise dans la narration musicale. Ces objets s'organisent selon des regroupements à plus grande échelle appelés systèmes d'implication qui, d'un point de vue cognitif, constitue l'ossature de processus d'attente, de réalisation ou de déni et structurent ainsi le discours musical.

De nombreux auteurs placent le phénomène d'attente au centre du processus musical. A ce titre, l'ouvrage de Huron intitulé *Sweet Anticipation: Music and the psychology of Expectation* (Huron, 2006) est particulièrement riche et couvre de nombreux versants de la question, qu'il s'agisse des volets culturels ou cognitifs de la question.

L'aspect culturel de la notion d'attente en musique est fortement lié à la question des conventions de composition qui sont au centre de la conception de structure musicale au sens Schenkerien (Schenker, 1935) et de la notion de cadence. Mais celle-ci est finalement assez délicate à définir (Neuwirth, 2015) surtout si l'on essaye d'englober de nombreux genres musicaux, non-classiques.

L'approche cognitive considère la notion de structure davantage sous l'angle des relations qu'entretiennent les éléments musicaux les uns avec les autres et comment leur enchaînement peut induire un effet conclusif dans la narration musicale hors de toute codification a priori.

C'est le parti pris adopté par Narmour dans (Narmour, 2000) pour qui les formes musicales sont régies par des phénomènes de projections cognitives qui sont tantôt réalisées, tantôt déniées.

Comme établi dans (Bimbot, 2016) le modèle Système & Contraste utilisé dans cette thèse constitue une généralisation du modèle de Narmour et se situe donc dans la catégorie des modèles de structure à fondements cognitifs. L'intérêt de cette approche est qu'elle peut se déployer à plusieurs échelles de temps simultanément et permettre de ce fait une description multi-échelle de la musique, y compris vis-à-vis de l'organisation interne des segments musicaux.

On considère donc que plus un motif présente des similarités avec un ou plusieurs motifs dans la section musicale à laquelle il appartient, plus ce motif sera considéré comme structurant ; et d'autant plus lorsque les couples de motifs sont placés dans des positions métriques homologues, à une échelle donnée.

Ainsi, il apparaît pertinent de chercher à définir une mesure qui exploite les relations de similarité au sein d'un passage musical pour en estimer sa complexité (c'est-à-dire le « contraire » de sa redondance). L'hypothèse étant que la capacité à regrouper un ensemble d'éléments musicaux dont la complexité conjointe est faible est un indice de leur appartenance probable à une même section musicale.

Un tel critère peut être mis en relation avec le concept de complexité de Kolmogorov. En supposant que deux éléments similaires sont reliés par une transformation « simple », c'est-à-dire de faible coût, la redondance du passage apparaît alors comme une fonction des coûts de transformation entre ses différents éléments.

Dans une approche causale, telle que celle que nous allons utiliser, nous calculerons la complexité d'un élément uniquement à partir de son historique d'éléments antécédents. En outre, les similarités entre éléments apparaissant à plusieurs échelles de temps simultanément, le critère de complexité intègrera ce point de vue multi-échelle.

La logique multi-échelle nous paraît essentielle, et favoriser les redondances apparaissant à des positions métriques homologues à plusieurs échelles de temps revient à faire l'hypothèse que leur synchronisation avec les différentes échelles métriques et hyper-métriques renforce leur impact structurant, du fait de leur concordance avec les cycles du « temps musical ».

1.1.4 Choix du style et éléments musicaux

La musique a été pratiquée de tout temps par la plupart des civilisations. C'est une des raisons pour lesquelles, il existe une quantité probablement indénombrable de styles et de genres musicaux. Dans cette étude, nous nous concentrons sur la musique tonale, occidentale et actuelle.

Même après cette première restriction, il demeure de nombreux genres musicaux qui ressortissent de cette catégorie... C'est pourquoi ce travail se focalise plus précisément sur des morceaux de musique Pop, un genre musical dont la structure interne des sections est généralement assez régulière.

L'idée n'est pas de considérer que la musique Pop est représentative de l'ensemble des genres musicaux. Il s'agit plutôt de tester un certain nombre de concepts, d'hypothèses et d'approches sur ce type de musique car il semble bien adapté aux modèles considérés, de qui permettra d'évaluer leur adéquation mutuelle.

En pratique, nous allons mener la présente étude sur le corpus RWC POP, composé de 100 chansons de musique Pop produites par le Japan's National Institute of Advanced Industrial Science and Technology (AIST). Ce corpus est bien connu dans le domaine du MIR et a déjà fait l'objet de multiples études.

Les travaux de cette thèse portent sur la structure interne des sections musicales, mais dans nos expériences nous restreindrons notre matériau expérimental dans 2 directions :

- La dimension musicale considérée sera la ligne mélodique, issue des versions MIDI des chansons de RWC POP.
- Les sections étudiées seront exclusivement les refrains des chansons, de 8 mesures en signature 4/4, ce sous-ensemble constituant 52 extraits musicaux.

Il existe de multiples conceptions du rôle de la mélodie dans la musique, et variables selon les genres musicaux et le rôle assigné aux autres dimensions musicales. Par exemple, dans une musique à base rythmique, la mélodie peut n'avoir qu'un rôle très mineur, voire être tout simplement absente. Mais dans de nombreux morceaux populaires traditionnels, notamment les chansons, elle joue un rôle central, et participe grandement à leur mémorisation et à leur appropriation vocale.

Dans l'immense majorité des extraits de notre corpus, les schémas mélodiques sont raisonnablement riches, et du fait de leur organisation carrée, les répétitions observées dans ces schémas les rendent facilement appréhendables cognitivement même en l'absence de support harmonique.

Nous répétons ici encore que l'ensemble des contraintes portant sur ce corpus vise à permettre d'étudier, dans un premier temps, les propriétés des approches à l'intérieur d'un périmètre bien délimité, en ayant conscience qu'il ne faut pas en tirer des conclusions par extrapolation. Mais les restrictions listées ci-dessus permettent de traiter un flux d'information unidimensionnel, bien identifié et dont la « carrure » est adaptée au modèle polytopique régulier présenté ci-après.

1.1.5 Critère et méthodologie

Comme évoqué brièvement dans la section précédente, un critère choisi pour caractériser le rôle structurant d'un motif musical dans un passage s'appuie sur une mesure de sa complexité, relative à son historique dans le passage considéré. Cette complexité se comprend comme (une approximation de) la quantité d'information nécessaire pour décrire ou construire un objet (Durand, 2007) et dans le cadre de cette thèse, il s'agira d'une complexité algorithmique, c'est-à-dire une approximation de la complexité du « programme » nécessaire pour expliquer le motif musical par rapport à son contexte. Cette approche se rattache donc à la notion de complexité de Kolmogorov, par opposition à l'information de Shannon.

Par ailleurs, les similarités ne sont pas supposées réparties aléatoirement dans une section musicale, ou tout du moins, selon nos hypothèses, leurs positions relatives impactent leur rôle dans la perception de la structure. En ce sens, une étape importante de cette étude est donc de définir de façon adéquate le contexte (ou l'historique) associé à chaque élément et nous nous baserons sur le modèle polytopique.

Une façon de vérifier la pertinence des hypothèses concernant l'identification de la structure par les similarités organisées selon le modèle polytopique est de mettre en œuvre des scénarios de recompositions musicales selon un ensemble particulier de fonctions de permutations qui préservent la complexité de la séquence au sens du critère de complexité.

Ces recompositions consistent à réorganiser ses motifs structurants dans un ordre différent de l'ordre séquentiel de départ. On obtient alors une nouvelle variation de la section musicale d'origine. Les permutations choisies pour générer ces recompositions conservent toutefois les propriétés structurelles liées à nos hypothèses de travail car elles intervertissent les éléments en positions métriques homologues sur le graphe polytopique support qui reflète ces relations d'homologie, et ce à différentes échelles temporelles.

Il n'est pas possible de prouver que toute composition musicale est construite selon un principe d'organisation multi-échelle polytopique (et ce n'est d'ailleurs probablement pas le cas !). Cependant, on peut évaluer dans quelle mesure un contenu musical qui vérifie ces propriétés reste cohérent vis-à-vis de ce qu'un panel d'auditeurs va considérer comme étant un contenu musical acceptable. Ceci s'effectue par une série de tests perceptifs où l'on demande à un panel d'auditeur de se prononcer sur l'acceptabilité de différents contenus musicaux recomposés.

Ainsi, on se propose d'évaluer à quel point la répartition de la redondance d'information multi-échelle est importante pour qu'un auditeur considère qu'un contenu musical est plus ou moins « valide » en vertu de principes structurels sous-jacents.

1.2 Système d'implication et modèle Système & Contraste (S&C)

1.2.1 Système d'implication

Le principe d'implication consiste à dire que lorsqu'un auditeur perçoit un élément musical, cela déclenche un processus cognitif d'attente qui place l'auditeur dans l'expectative quant à la similarité ou la dissimilarité de tout ou partie de l'élément musical à venir. Dans sa forme la plus simple, l'implication peut se formuler comme suit :

$$A \rightarrow A \quad (1.1)$$

Ce qui s'interprète comme « A implique A », cette implication pouvant être « réalisée » si l'élément suivant est à nouveau un A, ou bien « déniée » dans le cas où l'élément suivant est différent.

Comme expliqué par le principe d'implication de Narmour (Narmour, 1992), ce phénomène est rendu plus intense en présence d'éléments qui se répètent. Plus spécifiquement, la répétition de deux éléments similaires A induit encore plus l'attente d'un troisième élément similaire A, alors que deux éléments différents A et B provoquent (d'après Narmour) l'attente d'un troisième élément différent C.

$$A + A \rightarrow A \quad (1.2)$$

$$A + B \rightarrow C \quad (1.3)$$

Les principes de Narmour peuvent être compris comme un modèle cognitif basé sur la théorie de la Gestalt, comme mentionné dans (Narmour, Music expectation by cognitive rule-mapping, 2000).

Ce principe d'implication a été étendu par Bimbot et al. (Bimbot, 2016) sous la dénomination de modèle Système & Contraste, pour des séquences de 4 éléments, et s'écrit, en toute généralité :

$$A + f(A) + g(A) \rightarrow g(f(A)) \quad (1.4)$$

Où f et g sont des fonctions ou des relations qui expriment la transformation qui relie les éléments successifs, en référence au premier élément de la séquence, A , désigné par *primer*. Dans la plupart des cas, le 4^{ème} élément n'est pas $g(f(A))$, mais un élément qui s'y apparente plus ou moins, via une transformation supplémentaire dite « *contraste* » qui vient décrire la disparité entre l'attente et la réalisation du 4^{ème} élément.

On notera deux cas particuliers de la relation d'implication du modèle S&C :

$$A + B + A \rightarrow B \quad (1.5)$$

$$A + A + B \rightarrow B \quad (1.6)$$

qui sont à la base de deux schémas structurels très communs, notamment en musique Pop, *ABAC* et *AABC* dans lesquels l'attente induite par les 3 premiers éléments est rompue par le quatrième. Ces deux formes prototypiques se rattachent respectivement aux formes *période* et *sentence*, d'un point de vue musicologique. En fait, un nombre important de schémas musicaux suivent ce principe, et ce dans de multiples dimensions musicales : rythme, mélodie, harmonie, et souvent dans bien d'autres aspects de la composition du contenu considéré.

Comme développé dans (Bimbot, 2016) le principe du modèle S&C rend compte de systèmes carrés d'analogies partielles résultant de mécanismes d'attentes qui relient deux échelles de temps différentes simultanément. L'échelle temporelle correspondant à l'occurrence de l'élément suivant, et l'échelle temporelle correspondant à l'intervalle entre le premier élément et le troisième. C'est grâce à cette organisation sous forme « carrée » que peut se développer un modèle d'attente multi-échelle.

1.2.2 Système & Contraste

Dans sa version canonique, le modèle Système & Contraste est un cadre permettant de décrire les relations d'un groupe de quatre éléments $\{x_1; x_2; x_3; x_4\}$ formant un système.

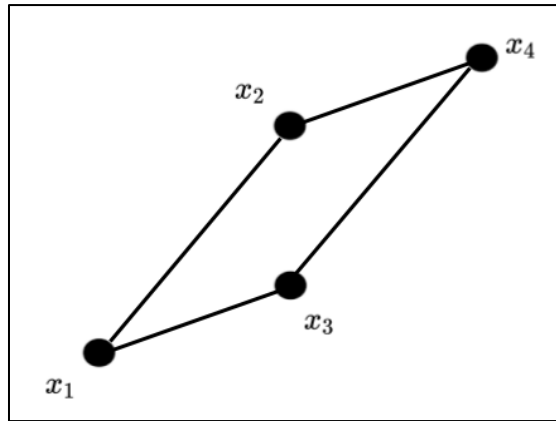


Figure 1: Forme carrée du modèle Système & Contraste

Ces quatre éléments peuvent être représentés en « carré » comme illustré (en perspective) sur la figure, cette disposition mettant en évidence le support de relations d'analogies entre les différents éléments du système.

Si le système n'est pas contrastif, des relations d'analogie complète sont observées, lesquelles s'expriment par :

$$x_2 \text{ est à } x_1 \text{ ce que } x_4 \text{ est à } x_3$$

$$\text{et/ou } x_3 \text{ est à } x_1 \text{ ce que } x_4 \text{ est à } x_2$$

Cette vision d'un groupe d'élément en tant que système permet de conférer à ses premiers termes un rôle de support à un mécanisme d'attente.

Par exemple, si l'on considère les systèmes ci-dessous, on peut aisément projeter l'élément manquant par un mécanisme d'implication purement « logique » :

A	B	2	4	∩	↓
E	.	8	.	∪	.

Figure 2 : Systèmes carrés d'implication

Dans un système complet, l'élément réellement observé en 4^{ème} position peut alors être apprécié cognitivement par rapport à l'élément « projeté » (ou impliqué), la différence entre les deux étant désignée par le terme « contraste ».

Ainsi, l'ensemble des éléments d'un système carré peut être décrit par (1) l'élément x_1 appelé le *primer* du système (2) deux transformations f et g modélisant respectivement les relations entre les éléments des couples $(x_1; x_2)$ et $(x_1; x_3)$ et (3) la fonction de *contraste* γ qui vient décrire la disparité entre le 4^{ème} élément du système et l'élément virtuel qui résulterait d'une pure relation d'analogie :

$$\begin{aligned} x_2 &= f(x_1); \\ x_3 &= g(x_1); \\ \text{et } x_4 &= \gamma(f(g(x_1))) \end{aligned} \tag{1.7}$$

Le triplet (x_1, f, g) est désigné système porteur. Lorsque les éléments du système possèdent plusieurs propriétés, il peut arriver que seul un sous-ensemble d'entre elles soit affecté par le contraste.

Il existe des extensions de la forme carrée du modèle Système & Contraste, permettant de modéliser des séquences de 3, 5, 6, ... éléments en « déformant » le système canonique. Ces variantes sont détaillées dans (Bimbot, 2016) et permettent de rendre compte de structures musicales variées.

Dans les travaux de cette thèse, nous utilisons le modèle Système & Contraste à plusieurs échelles temporelles à l'intérieur d'une structure polytopique régulière où chaque face comporte 4 éléments formant un système.

1.3 Généralisation multi-dimensionnelle et graphes polytopiques

Le modèle Système & Contraste tel qu'introduit ci-dessus, fait intervenir 4 éléments, soit indirectement 2 échelles de temps.

À partir de la forme carrée du modèle Système & Contraste, les études menées par Bimbot, Deruty, et complétées par Guichaoua, ont montré qu'il était possible de généraliser la forme carrée du modèle S&C en incorporant des niveaux croissants d'analogie (Deruty, 2013).

Ainsi, si l'on met en correspondance 2 systèmes de 4 éléments, on crée un nouveau système formé de leurs 4 arêtes, lequel donne lieu à une structure de 8 éléments à support cubique, mettant en jeu 3 échelles de temps. Puis si l'on reproduit ce principe, on obtient un système de 16 éléments supporté par un tesseract. Et ainsi de suite, avec pour résultat général un n-cube (cf. Figure).

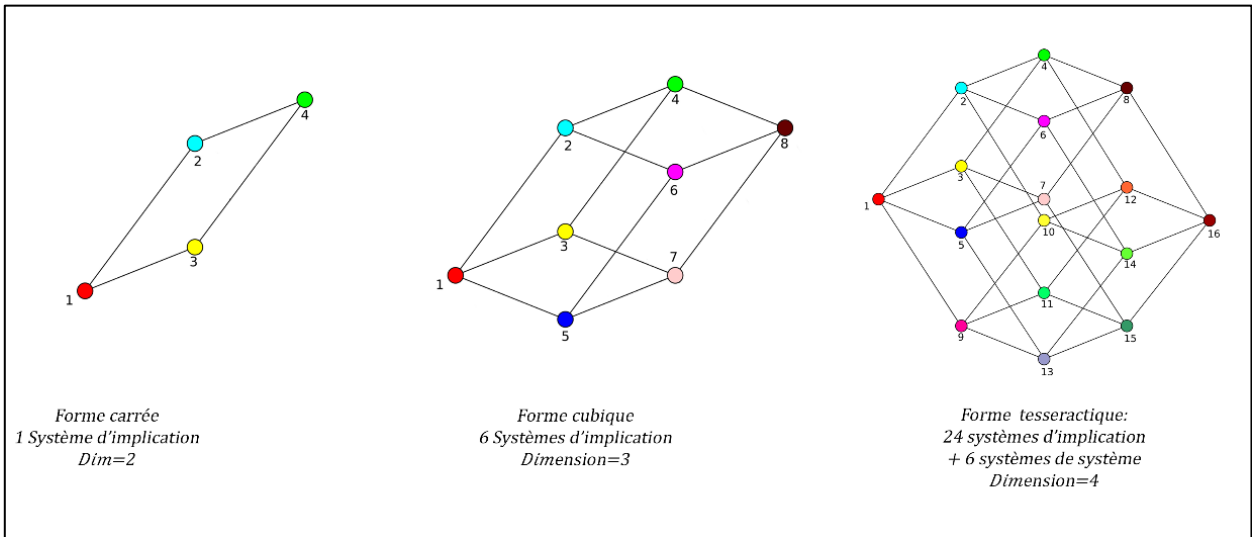


Figure 3: Généralisation multi-échelle des systèmes d'implications

Généralisation multi échelle des systèmes d'implications

On remarque que lorsque l'on passe d'un niveau au niveau immédiatement supérieur, de nouveaux systèmes apparaissent.

Par exemple, lors du passage de la forme carrée à la forme cubique, 6 systèmes d'éléments se font jour, correspondant à chacune des faces du cube, ainsi que 3 systèmes d'arêtes. Pour le tesseract, on aboutit à 24 systèmes d'éléments (24 faces), 24 systèmes d'arêtes (8 cubes x 3 systèmes/cube) et 6 systèmes de faces, ces derniers constituant de fait des systèmes carrés.

Plus généralement, ce principe peut être généralisé à des systèmes n -cubiques où n est le nombre d'échelles considérées simultanément.

Dans cette thèse, on va utiliser la généralisation du modèle Système & Contraste en dimension 4 qui permettra ainsi de rendre compte des relations structurelles à 4 échelles temporelles différentes, comme cela est schématisé sur la Figure 4. Le polytope sera donc un tesseract (4-cube).

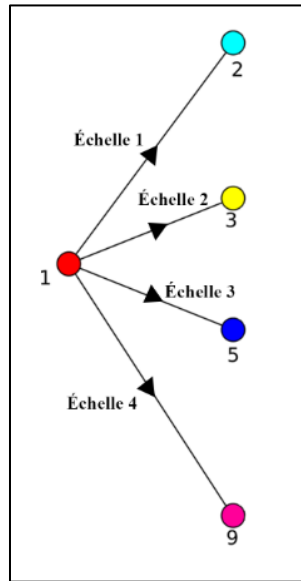


Figure 4: Présentations des relations structurelles entre éléments à différentes échelles temporelles.

On remarque sur la figure que l'orientation des arêtes du tesseract nous renseigne sur l'échelle temporelle à laquelle a lieu la relation entre deux éléments reliés l'un à l'autre.

Le tesseract est un polytope régulier dont les faces sont des carrés. C'est pourquoi, notre étude va porter sur des sections musicales compatibles avec cette géométrie et que nous avons fait le choix, comme cela est évoqué plus haut, de travailler sur des refrains de musique POP d'une durée de 8 mesures. Ainsi, chaque sommet du tesseract sera porteur d'un motif musical d'une durée de $\frac{1}{2}$ mesure de métrique quaternaire.

Ce principe peut être aisément généralisé à d'autres types de métriques. Par exemple dans le cas d'une section à base ternaire, on peut travailler avec un polytope à faces triangulaires. Toutefois, ces considérations sortent des travaux de la présente thèse et pourront être explorées dans des travaux futurs.

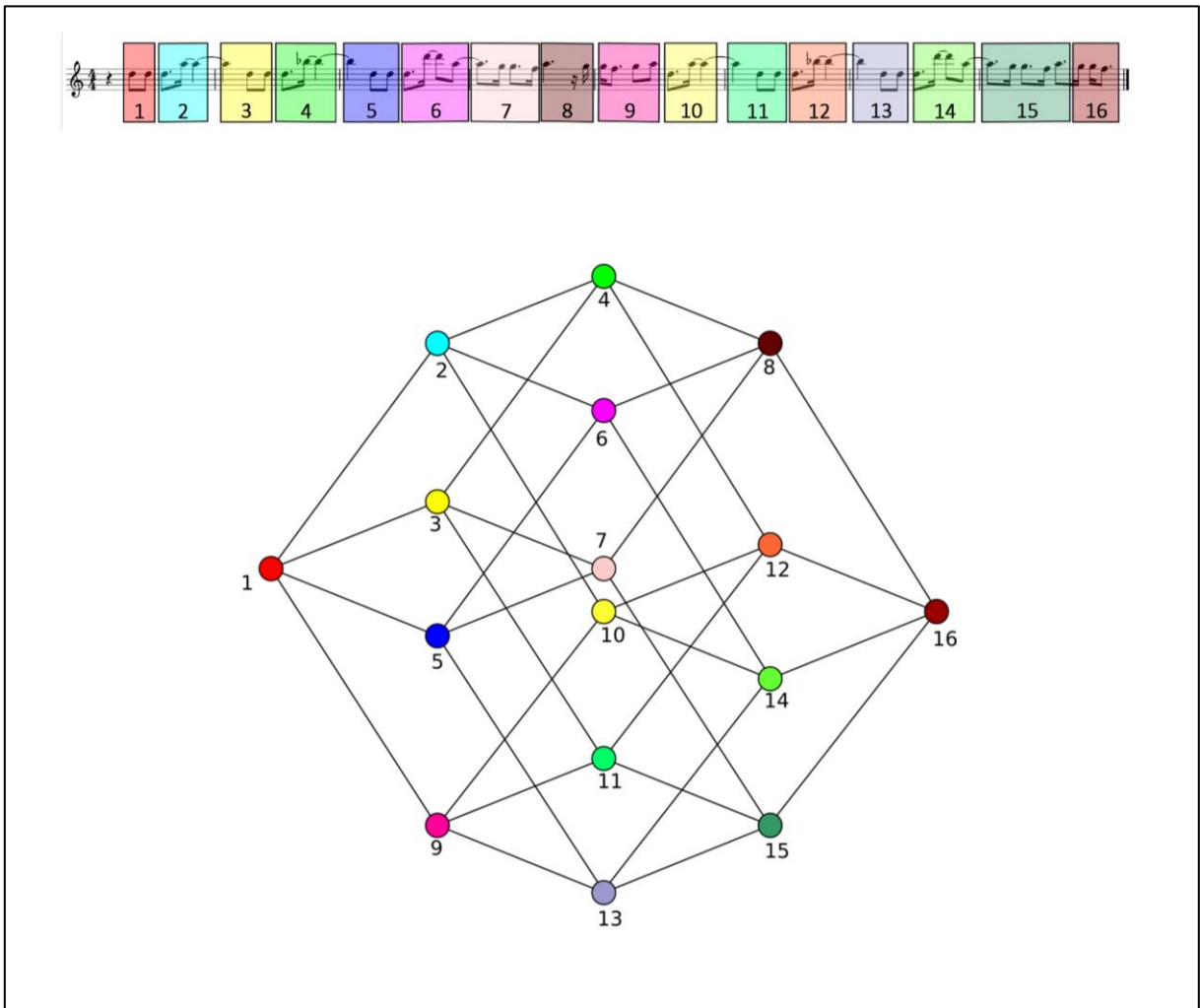


Figure 5: Passage de la section musicale en vue séquentielle à une vue polytopique de dimension 4.

La structure polytopique peut être considérée comme étant le support d'un graphe déterminant les relations considérées comme étant d'intérêt prépondérant dans une section musicale. Cette conception a été étudiée dans les travaux de Louboutin (Louboutin & Bimbot, Polytopic Graph of Latent Relations: A Multiscale Structure Model for Music Segments, 2017) qui utilise notamment ce formalisme dans le but d'analyser et de générer des suites d'accords.

Dans ses travaux, Guichaoua (Guichaoua, 2017) utilise également des graphes polytopiques pour segmenter un morceau de musique en sections.

Dans les travaux de la présente thèse, ce support de graphe va être utilisé afin de rendre compte visuellement d'une section musicale sous une forme délinéarisée permettant de caractériser les relations développées à plusieurs échelles temporelles entre les différents motifs constituant cette section dans le but de recomposer la section concernée.

Chaque motif de la section musicale étudiée est relié à un ensemble de motifs antécédents constituant son historique. Les arêtes et/ou les faces définies par ce support sont ainsi utilisées afin de définir cet historique, et indiquer quels sont les motifs antécédents les plus pertinents à considérer.

Un des points clefs de cette thèse réside en ce qu'il existe plusieurs façons de parcourir les sommets de ce polytope support. Ce processus a pour effet de modifier l'ordonnancement des motifs de la section musicale et ainsi de produire une variation de cette section, que l'on appelle une recomposition polytopique, dès lors que le nouvel ordonnancement respecte certaines propriétés.

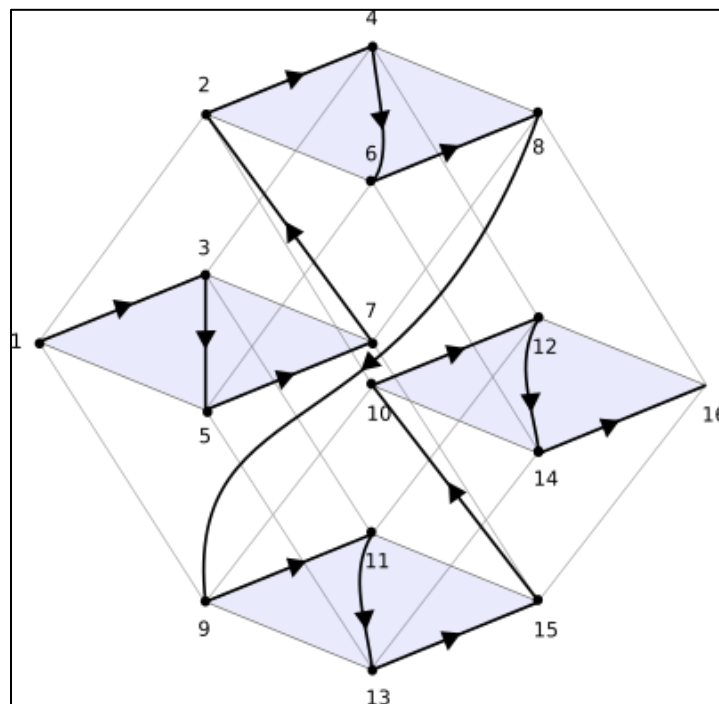


Figure 6: Exemples de parcours du polytope support hypercube dans un ordre non séquentiel.

1.4 Variations et Recompositions Polytopiques

1.4.1 Discussion par rapport aux variations existantes

La variation est une technique de composition qui consiste à modifier un ou plusieurs paramètres musicaux d'un matériau de départ et le transformer dans certaines de ses dimensions musicales, au point parfois de le rendre quasiment méconnaissable.

Cette pratique est fondamentale dans la composition musicale écrite ainsi que dans les musiques de tradition orale.

Traditionnellement, la variation s'obtient par l'altération de la mélodie, du rythme de l'harmonie ou encore de l'instrumentation.

Voici quelques exemples de la pratique des variations dans l'histoire de la musique, issue de (Sadie, 2001) et (Austin, 1966).

À la Renaissance, on voit apparaître des genres musicaux entièrement basés sur la technique de la variation. On répète une basse toujours identique en ostinato, ce qui permet aux voix supérieures de varier le propos mélodique à l'infini.

De nombreux genres musicaux sont fondés sur la variation à l'époque baroque : les pièces des basses obstinées, les chacones, passacaille, les grounds en Angleterre, mais aussi les danses variées accompagnées de leurs doubles qui propose une alternative ornementale à la danse initiale.

Dans le répertoire spécifique de l'orgue, il y a le choral varié et le thème varié où Jean Sébastien Bach a excellé dans les variations Goldberg pour clavecin publié en 1741. Cette pièce est constituée de 30 variations d'une aria fondée sur une basse italienne très répandue à l'époque.

À l'époque classique, le principe des variations devient un élément constitutif des sonates et de tous les genres de musique de chambre. Joseph Haydn aime faire appel au thème des variations dans ses mouvements lents et Mozart dans ses finals.

En 1823, Beethoven présente une œuvre majeure de 33 variations dans son Opus 120, il s'agit des « Variations Diabelli » qui restent aujourd'hui un chef-d'œuvre dans la technique de la variation.

Pour varier habituellement une mélodie on peut toucher à tous les paramètres de la musique.

Par exemple, Debussy dans son prélude à l'après-midi d'un faune varie le thème de la flute sans modifier à une seule note de ce thème, mais en lui confiant un accompagnement harmonique différent à chacune de ses répétitions.

Dans le Boléro de Ravel, la mélodie est toujours la même, mais à chaque répétition, il change d'instrument. La mélodie est donc ici variée par son timbre.

Michaël Levinas, dans Étude n°2 : Variations sur une seule note, varie notamment dans les modes d'attaque, mais aussi sur les modes de jeu en altérant les cordes par un contrôle de la main directement sur les cordes, ce qui crée des variations de la résonance du son.

Dans ses travaux de thèse, Louis Bigo (Louis Bigo, 2014) utilise un graphe support représentant l'espace tonal basé sur le Tonnetz, où une séquence musicale y est représentée par une trajectoire dans ce graphe. Des variations musicales sont apportées à cette séquence à partir d'opérations de translations, rotations ou de symétrie de son support géométrique.

Ce principe est assez proche de celui que nous allons développer dans ces travaux. On définit un ensemble de variation d'une séquence musicale à partir d'opérations sur son graphe polytopique support, mais dans le cas présent les transformations considérées sont des permutations des nœuds du graphe. On les appelle ces variations des recompositions polytopiques et elles sont présentées en détail ci-après.

1.4.2 Recompositions Polytopiques

Alors que des modifications de l'ordre dans lequel apparaissent les sections musicales sont couramment utilisées pour produire des « remix », les variations dans l'ordonnancement temporel de la musique à l'intérieur d'une section donnée sont peu communes.

Au même titre que les autres variations évoquées dans ce chapitre, les variations de l'ordre temporel que nous allons étudier veillent à préserver certaines caractéristiques topologiques de l'ordonnancement du contenu.

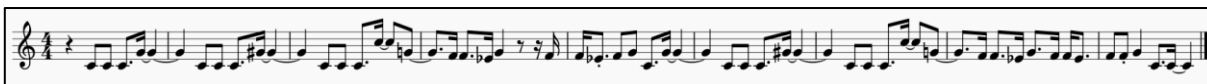


Figure 7 : Mélodie du refrain de Adele – Skyfall.

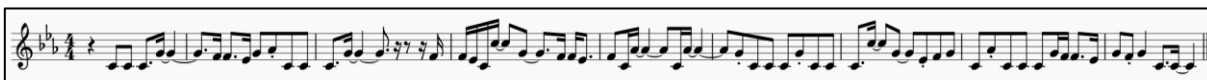


Figure 8 : Recomposition aléatoire de la mélodie du refrain de Adele - Skyfall.

Sur la Figure 8, la mélodie de la chanson Skyfall de Adèle est recomposée aléatoirement à l'échelle des motifs (d'une demi-mesure). Par « aléatoire » nous entendons que les motifs sont réordonnés sans tenir compte d'une quelconque stratégie ou contrainte quant à leur recombinaison.

On peut alternativement choisir de les grouper 2 par 2, 4 par 4 ou définir différents systèmes de groupements tenant compte d'une logique structurelle prédéfinie. C'est cette dernière stratégie qui sera développée dans ces travaux.

La stratégie de recomposition polytopique consiste à appliquer une permutation des échelles qui a pour effet d'altérer de façon déterministe l'ordre temporel des motifs musicaux de la section d'origine. En conséquence, les variations polytopiques préservent un certain nombre de propriétés structurelles.

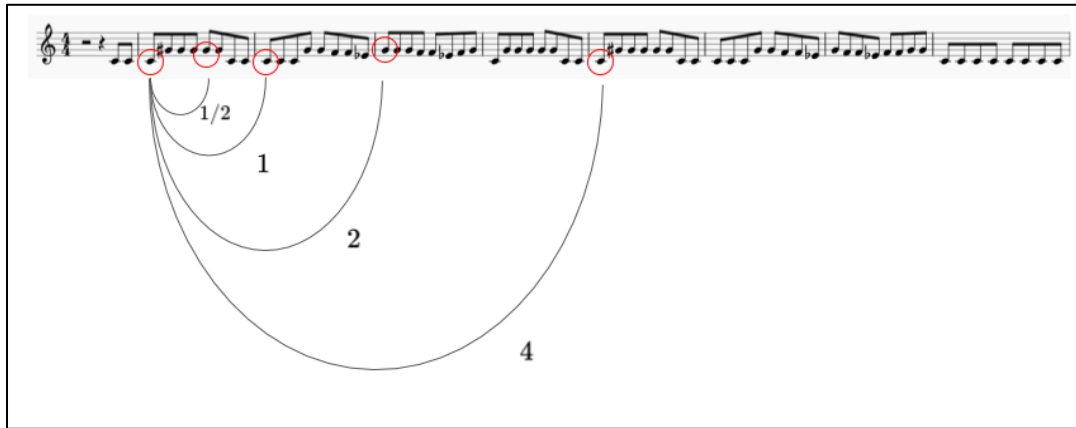


Figure 9 : Illustration des notes en position métrique homologue à la note n°1 à différentes échelles métriques (1/2 mesure, 1 mesure, 2 mesures et 4 mesures).

Ainsi, en intervertissant les échelles dans le polytope, on effectue une permutation des éléments musicaux situés en position métrique homologue, et dans le tesseract, ces échelles de temps correspondent à la $\frac{1}{2}$ mesure, la mesure, 2 mesures et 4 mesures.

Sur la figure 9, on représente par des notes entourées celles qui se trouvent en positions métriques homologues à la 1ère note, aux 4 échelles temporelles représentées dans un tesseract. Plus généralement, chaque nœud dans le tesseract est en relation avec 4 autres nœuds qui représentent les éléments en position métrique homologue, soit en tant qu'antécédents, soit en tant que successeurs.

Les notes, ou groupements de notes ainsi associés entretiennent des relations privilégiées qui sont exploitées dans l'étude multi-échelle de la structure et les nœuds antécédents d'un nœud donné constituent, comme on l'a évoqué précédemment, un historique privilégié dudit nœud dans le modèle polytopique. On rappelle d'ailleurs que les faces du polytope forment des systèmes d'implications comme montrés sur la figure ci-après.

On caractérise ces variations polytopiques par des fonctions bijectives appliquées aux index des motifs structurants d'une section musicale. On considère en particulier, un sous ensemble de ces fonctions qui ont la particularité de préserver les systèmes de relations identifiées sur les faces du polytopes support, mais de les faire intervenir à des échelles temporelles différentes. Ces fonctions sont nommées les Primer Preserving Permutations (PPP).

Finalement, appliquer une permutation PPP sur le graphe polytopique d'une section musicale permet d'obtenir une nouvelle version de celle-ci où les motifs seront placés et joués dans un ordre différent.

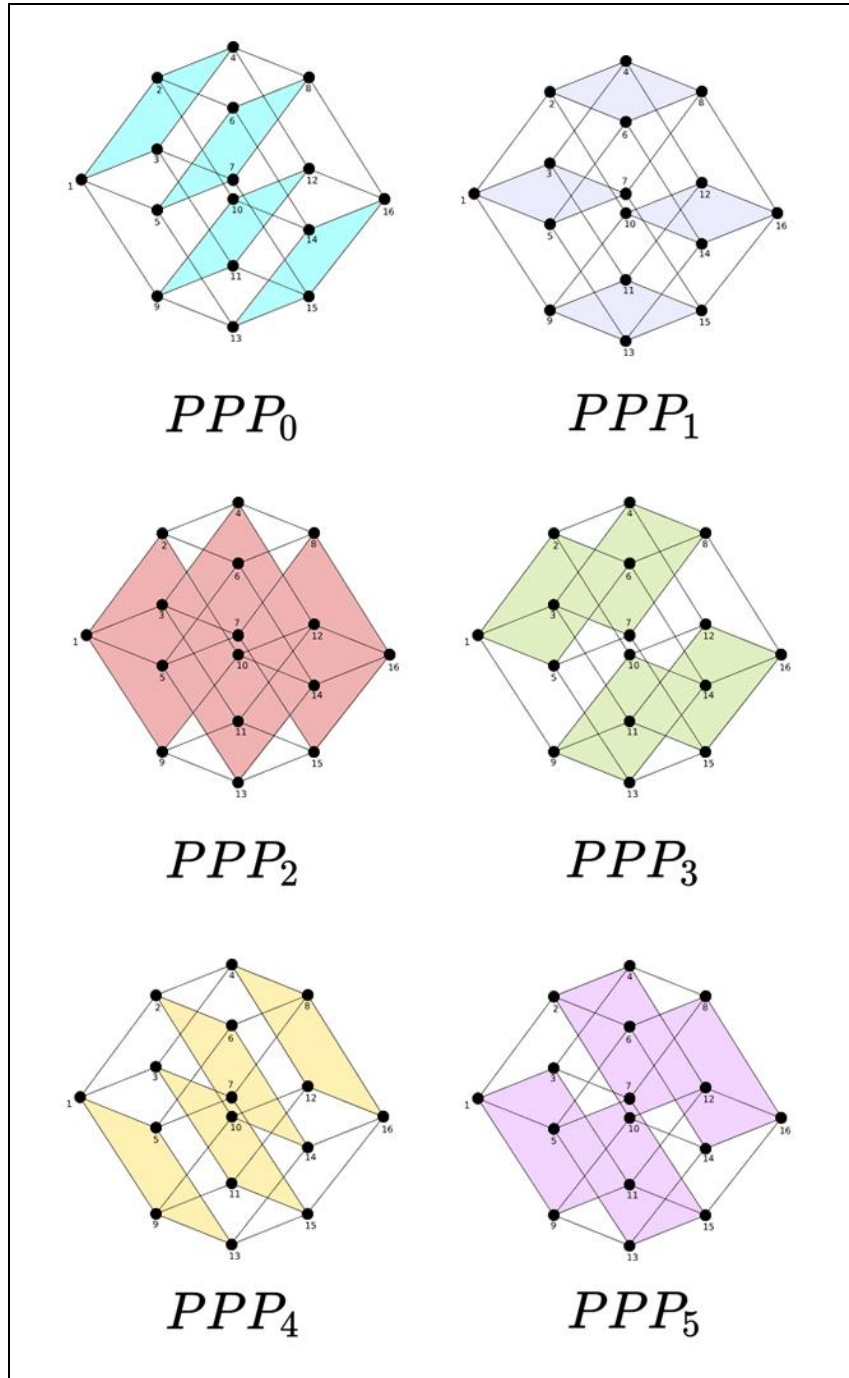


Figure 10: Représentation des 6 systèmes de face parallèles sur le tesseract et leur permutation PPP correspondante.

1.5 Théorie de l'Information et complexité

La communauté MIR (Music Information Retrieval) s'intéresse de près à la Théorie de l'information, notamment pour des tâches d'apprentissage. En effet, les techniques liées à la théorie de l'information permettent de traiter des grands jeux de donnée et de mettre en place des méthodes statistiques comme les modèles de Markov. Les travaux de (Raphael, 2002) utilisent par exemple un modèle de chaîne de Markov cachée pour des tâches de transcriptions automatiques de musique jouée au piano.

Ces approches sont fortement liées à la théorie de Shannon (Shanon, 1948) dans laquelle il décrit que les données observées peuvent être considérées comme la sortie d'un système de communication et leur coût d'encodage lié à leur probabilité d'émission.

Dans le domaine de la recherche d'information musicale, les modèles probabilistes servent à estimer la distribution des événements musicaux, comme les notes, le rythme, les accords... sur la base de leurs fréquences d'apparition dans un corpus.

Dans la théorie de Shannon, un événement musical est la réalisation d'une variable aléatoire qui suit une distribution de probabilité estimée selon les fréquences d'apparitions de ces événements dans un corpus. Dans ce cas, la complexité de l'événement musical repose sur la caractérisation de la variable aléatoire par un processus statistique capable de générer cet événement avec la même probabilité que celle mesurée.

L'information issue de la théorie de Kolmogorov peut être vue comme une analyse complémentaire à celle de Shannon (Meredith, 2012) (Bimbot, 2016).

D'après celle-ci, on décrit un objet musical à partir du programme le plus court capable de générer cet objet. La complexité de Kolmogorov est la distance de description minimale (MDL) de l'objet musical, c'est-à-dire, la taille (en nombre de bits) du programme le plus court qui ne prend pas de données d'entrées et qui est capable de retourner l'objet musical en sortie.

Il est très complexe de prouver qu'une description algorithmique donnée est la plus petite possible. C'est pourquoi cette description doit être estimée et approximée. La meilleure méthode pour approximer la complexité de Kolmogorov consiste à mesurer la compressibilité de l'objet musicale au sens d'une famille de programmes pré-déterminés par exemple par des combinaisons de transformations simples entre parties de l'objet et ainsi approximer son niveau de redondance interne (ou compressibilité).

Par exemple, si l'objet musical est une section dont les parties sont ses motifs, un motif donné contribuera d'autant moins à la complexité de la section, s'il présente de fortes similarités avec les autres motifs présents.

Les travaux de cette thèse sont principalement basés sur l'approche algorithmique pour calculer la complexité d'une section musicale par l'intermédiaire de la description de ses motifs à partir de leur « historique » au sein de la section.

1.6 Méthodologie

Dans les parties précédentes, un certain nombre d'hypothèses concernant la structure musicale ont été formulées. On considère, par exemple, que la structure d'une narration musicale se développe à plusieurs échelles de temps différentes et simultanément.

Cette hypothèse a pour conséquence de représenter une section musicale de façon délinéarisée à l'aide de son graphe polytopique support et de mettre en évidence ses relations multi-échelle, chaque sommet de ce support polytopique correspondant à un motif structurant de la section musicale considérée. En lien avec la théorie de Narmour, plus un motif musical est similaire à d'autres éléments de la section, plus il est considéré comme structurant, car il participe à un processus d'attente.

La tâche de segmentation pour instancier ces motifs est au cœur des travaux de cette thèse et plusieurs méthodes sont considérées, qui sont développées dans le chapitre suivant.

Une fois les motifs structurants issus de la segmentation, on considère qu'il est possible d'appliquer certaines fonctions de recompositions permutant les sommets du support polytopique et ainsi générer de nouvelles versions de la section musicale de base, tout en conservant certaines propriétés de structure.

Ces hypothèses demandent à être évaluées et deux problématiques principales se posent :

- Comment caractériser la ou les segmentations optimales d'une phrase musicale de façon à ce que chaque motif soit un objet musical cohérent et structurant ? Si les motifs structurants ne sont pas bien délimités par la segmentation, les recompositions ne pourront pas être satisfaisantes.
- Comment mesurer empiriquement ou objectivement l'admissibilité d'un contenu qui résulte d'une recomposition polytopique ?

En complément, le travail présenté dans la suite de cette thèse propose aussi un prototype d'outil créatif reposant sur les concepts développés à l'occasion de ces travaux.

1.6.1 Procédé de recomposition

En supposant qu'une section musicale résulte de la juxtaposition de motifs dont les frontières sont prédéfinies, le procédé de recomposition consiste à instancier des scénarios pour la génération des variations de la section musicale d'origine.

On rappelle que ces recompositions consistent à former une nouvelle version de la section musicale d'origine en faisant varier l'ordre temporel de ses motifs constituants.

Cette réorganisation temporelle peut être aléatoire, mais elle est peut également être régie par des contraintes, notamment des hypothèses de conservation des relations structurantes, énoncées dans les sections précédents.

Ainsi, en appliquant les principes du modèle S&C au tesseract, les fonctions de recompositions dites polytopiques réorganisent temporellement les motifs, tout en conservant, comme nous le détaillerons plus loin, le coût de description de la séquence originale. En d'autres termes, une recomposition polytopique laisse invariante la complexité de la séquence, au sens de la structure issue du modèle S&C.

1.6.2 Algorithmes de segmentation

Si la segmentation interne de la section en motifs n'est pas prédéfinie, il est nécessaire de la déterminer en délimitant les motifs structurants par optimisation d'un critère. En effet, la surface musicale des motifs n'est pas nécessairement synchronisée sur les débuts de mesure, comme c'est le cas, par exemple, pour les anacrouses.

Cette tâche peut être réalisée de plusieurs façons.

La segmentation de la section musicale en motifs peut être confiée à un expert du domaine capable de les délimiter manuellement, mais avec un coût de gestion des données important.

Cette approche est toutefois peu réaliste et un des buts des travaux de cette thèse est de définir une méthode permettant d'obtenir cette segmentation de façon automatique.

Cette tâche est donc réalisée par un algorithme qui teste plusieurs hypothèses de segmentation sur la section musicale et détermine celle qui est optimale. Ici intervient un critère de complexité qui sera détaillé dans le chapitre 2.

1.6.3 Caractérisation des variations polytopiques

Une fois générée une ou plusieurs variations polytopiques d'une section musicale, se pose la question d'évaluer la validité musicale de ce type de variation. Il ne s'agit pas tant de mesurer la validité en tant que telle, mais surtout d'évaluer si les recompositions polytopiques présentent un avantage en termes de cohérence musicale par rapport à des recompositions aléatoires. Ceci afin d'explorer le niveau de pertinence d'un critère de complexité qui est invariant par permutation d'échelles.

Il n'existe pas de critère objectif pour juger de la cohérence musicale d'un passage musical. C'est donc par le biais de jugements perceptifs par un panel d'auditeurs humains que nous allons évaluer cette propriété.

On va tout d'abord caractériser les propriétés topologiques inhérentes à différentes recompositions, à savoir, le nombre de discontinuités induites entre motifs recomposés par rapport à l'ordre séquentiel de la section de base. On considère également une quantité mesurant la distance temporelle maximale induite par la fonction de recomposition, ce que l'on appelle excursion entre motifs dans la section recomposée par rapport à la section originale.

1.6.4 Validation empirique

Les quantités discutées dans la partie précédente donnent des indicateurs a priori de l'acceptabilité musicale, basés sur des propriétés purement topologiques qui caractérisent certaines propriétés de déformation subies par le contenu musical d'origine. Mais elles sont purement topologiques et ne tiennent pas compte du contenu proprement dit.

L'appréciation d'un contenu musical étant quelque chose d'extrêmement subjectif, il n'existe guère d'autre solution pour évaluer la validité musicale d'un extrait, que de mettre en place des tests perceptifs pour évaluer empiriquement cette appréciation selon un panel d'auditeurs. En termes d'évaluation, on propose donc à une population d'auditeur d'évaluer subjectivement un certain nombre de recompositions, comprenant des morceaux originaux, des recompositions polytopiques et des recompositions aléatoires.

L'analyse de ces résultats tests permet de situer quantitativement les recompositions polytopiques par rapport à de passages originaux d'une part et des recompositions aléatoires d'autre part.

1.6.5 Conception d'un outil créatif

Les fondements développés dans ce travail peuvent également être mis à profit à des fins créatives. En effet, la vision délinéarisée de la structure musicale est une idée novatrice qu'il est intéressant de partager avec des compositeurs, afin qu'ils puissent en explorer le potentiel artistique et créatif dans un foisonnement d'idées nouvelles en termes de variations musicales.

Dans le cadre des travaux de cette thèse et en partenariat avec l'entreprise Kercodex, un prototype d'interface de création et de recomposition musicales a été développé et est disponible (sous forme de preuve de concept) pour le grand public. Cette interface se présente sous forme d'application mobile qui permet à n'importe quel utilisateur d'utiliser les fondements développés dans ces travaux afin de visualiser une section musicale de façon non linéaire et d'en générer plusieurs recompositions de façon automatique.

Chapitre 2

Concepts et Formalisme

2.1 Objets musicaux & échelles temporelles

Dans cette partie, nous précisons et formalisons l'ensemble des objets musicaux, ainsi que les différentes échelles temporelles qui sont à l'étude dans ces travaux.

La majeure partie des études liées à la structure musicale se situent à l'échelle temporelle d'un morceau entier, où sont étudiés les agencements entre thèmes, introductions, couplets, refrains, etc.

Dans cette thèse, on se situe à une échelle plus basse, qui permet l'analyse de l'organisation structurelle intrinsèque d'une section musicale, c'est-à-dire, dans le cas de la musique Pop, la construction interne d'un couplet, d'un refrain, d'un pont... Notre étude porte d'ailleurs plus particulièrement sur l'analyse des refrains, qui sont des sections qui jouent un rôle essentiel dans la structure globale d'une chanson.

Les musiques pops, sur lesquelles se focalise ce travail, correspondent par ailleurs à une grande majorité de la musique écoutée et produite aujourd'hui. Quoiqu'on pense du niveau de sophistication de ces musiques, elles constituent un sujet d'analyse pertinent au regard de leur fréquence d'usage, de leur rôle sociologique, de la masse et de la diversité de données disponibles.

Même s'il existe quelques exceptions, la ligne mélodique des refrains pops jouent un rôle essentiel dans la perception de la musicalité et participe de façon majeure à leur structure. Et comme nous l'avons évoqué précédemment, les refrains sont majoritairement composés de 8 mesures en signature temporelle 4/4 ce qui s'adapte bien au modèle polytopique par tesseract.

Pour situer musicologiquement le positionnement des échelles d'étude, la figure 11 vient décrire les différents niveaux d'expériences musicales selon (Snyder, 2000) : pas moins de 27 échelles qui vont du 10 millième de seconde, à un peu plus d'une heure !

Dans le présent travail, on manipule des objets musicaux intervenant dans les échelles « Melodic and rhythmic grouping » sur l'échelle de Snyder, correspondant à la mobilisation de mémoire à court terme. Par opposition aux échelles microscopiques (« Event ») et macroscopiques (« Form »), on pourrait désigner cette échelle d'étude, comme étant l'échelle *mésoscopique*.

Une des étapes essentielles dans ce travail consiste à repérer et instancier les motifs élémentaires structurants correspondant à l'échelle mésoscopique (donc intermédiaire) sur la table de Snyder.

	Events per second	Seconds per event
EVENT FUSION (early processing)	16,384	1/16,384
	8,192	1/8,192
	4,096	1/4,096
	2,048	1/2,048
	1,024	1/1,024
	512	1/512
	256	1/256
	128	1/128
	64	1/64
	32	1/32
MELODIC and RHYTHMIC GROUPING (short-term memory)	16	1/16
	8	1/8
	4	1/4
	2	1/2
	1	1
	1/2	2
	1/4	4
	1/8	8
FORM (long-term memory)	1/16	16
	1/32	32
	1/64	1 min 4 sec
	1/128	2 min 8 sec
	1/256	4 min 16 sec
	1/512	8 min 32 sec
	1/1,024	17 min 4 sec
	1/2,048	34 min 8 sec
1/4,096	1 hr 8 min 16 sec.	

Figure 11: Tables des niveaux d'expériences musicales extrait de (Snyder, 2000).

Sections

Dans ces travaux, les sections musicales sont segmentées en $z = 16$ motifs notés ainsi :

$$x_1, \dots, x_t, \dots, x_z \text{ avec } 1 \leq t \leq z = 16. \quad (2.1)$$

D'un point de vue perceptif, pour qu'un auditeur identifie un motif musical, harmonique, mélodique ou d'une autre nature, il est nécessaire que ce dernier ait une durée suffisante et ainsi pouvoir être identifié cognitivement en tant qu'élément musicalement intentionnel en soi.

Ainsi, si on manipule des fragments musicaux trop courts, l'auditeur ne sera pas en situation de les percevoir comme des entités musicales en tant que telles et les recompositions temporelles de ces éléments seront perçues comme erratiques.

De plus, il est important de tenir compte de la fréquence de « réactualisation harmonique » qui ne doit pas être trop rapide pour que les caractéristiques de la musique tonale restent préservées.

Ainsi, l'échelle temporelle de recomposition ne doit pas être trop fine si l'on vise à préserver des propriétés cohérentes avec celles de l'échelle mésoscopique.

À l'inverse, l'échelle temporelle des motifs ne doit pas être trop grande, au risque de limiter considérablement les possibilités de variations de l'ordre temporel de ces objets. Par exemple, recomposer des sections musicales de 8 mesures par groupes de 4 mesures offre clairement des possibilités limitées...

Motifs

Dans ces travaux, on fixe la taille d'un motif élémentaire à $\frac{1}{2}$ mesure.

Afin de pouvoir décrire les motifs eux-mêmes d'une façon suffisamment fine, on définit des éléments dits *granulaires*.

Un motif x_j est constitué de $m = 4$ granules.

$$x_j = u_1^{(j)} \dots u_i^{(j)} \dots u_m^{(j)} \text{ avec } 1 \leq i \leq m = 4 \quad (2.2)$$

Ces éléments constituent l'unité la plus fine de ces travaux avec une durée de $\frac{1}{8}$ de mesure.

Granules

Une section de 8 mesures est donc constituée de 64 granules définis ainsi :

$$u_1, \dots, u_b, \dots, u_g \text{ avec } 1 \leq b \leq g = 64 \quad (2.3)$$

L'instanciation de ces différents objets va être détaillée dans les sections suivantes et une table récapitulative des objets manipulés est présentée en page suivante.

Il est clair que le choix de ces différentes échelles est en partie arbitraire et que dans des études futures il est envisageable de s'en affranchir.

Mais ici les facteurs d'échelle sont placés dans un rapport 4 ce qui est propice à des formalismes mettant en jeu des systèmes carrés de type S&C.

Différentes échelles temporelles manipulées dans ces travaux

Granules

$\frac{1}{8}$ de mesure

U_b

A musical score in 4/4 time with a treble clef and a key signature of one flat. The score consists of eight measures. A red box highlights the fifth measure, which contains a quarter note G4 and an eighth note A4. Below the score, a diagram shows the measures numbered 1 to 8, with a vertical line connecting the fifth measure to the highlighted red box.

Motifs

$\frac{1}{2}$ de mesure

X_t

A musical score in 4/4 time with a treble clef and a key signature of one flat. The score consists of eight measures. A red box highlights the fifth measure, which contains a half note G4. Below the score, a diagram shows the measures numbered 1 to 8, with a vertical line connecting the fifth measure to the highlighted red box.

Systemes

2 mesures

S_k

A musical score in 4/4 time with a treble clef and a key signature of one flat. The score consists of eight measures. A yellow box highlights the fifth and sixth measures, which contain a half note G4. Below the score, a diagram shows the measures numbered 1 to 8, with a vertical line connecting the fifth measure to the highlighted yellow box.

Section Musicale

8 mesures

X

A musical score in 4/4 time with a treble clef and a key signature of one flat. The score consists of eight measures, each containing a quarter note. The notes are G4, A4, Bb4, C5, Bb4, A4, G4, and F4. Each measure is highlighted with a different color: red, green, yellow, purple, blue, orange, pink, and light blue. Below the score, a diagram shows the measures numbered 1 to 8, with a vertical line connecting the first measure to the first highlighted measure.

Figure 12: Table récapitulative des échelles manipulées dans ces travaux.

2.1.1 Forme réduite de l'information musicale

Dans ce travail, on manipule des sections musicales correspondant à des refrains de musique Pop de 8 mesures avec une signature temporelle 4/4.

Les analyses menées ici portent sur une représentation symbolique de la musique et on travaille donc sur la partition de ces extraits.

Afin d'analyser algorithmiquement les données présentes dans la partition d'une section, on manipule des fichiers au format MIDI, où chaque piste correspond à une représentation de type piano-roll des divers instruments de l'extrait étudié.

Plus précisément, on s'intéresse au développement et à l'organisation structurelle de la ligne mélodique.

Afin de pouvoir analyser cette structure par le biais de différents critères, il est nécessaire de poser des hypothèses de travail et de normaliser la représentation des données manipulées.

La section musicale X dans sa forme d'origine est donc être échantillonnée en un ensemble de 64 éléments granulaires $u_1 \dots, u_b, \dots, u_g$ ($1 \leq b \leq g = 64$), ce qui va produire une version réduite de cette même section.

$$X = u_1 \dots u_b \dots u_g \quad 1 \leq b \leq g = 64 \quad (2.4)$$

Les notes se produisant entre les instants d'échantillonnage sont volontairement omises de la version réduite, mais remplacée par la note active durant le début de l'échantillon.

On considère que le rythme de l'information mélodique *structurante* ne se produit pas à une échelle inférieure à la noire (1/4 de mesure en signature 4/4). Cette durée métrique conditionne donc la fréquence d'échantillonnage de l'information à un rythme double (c'est-à-dire la croche), en vertu du principe de Shannon. D'où des granules de 1/8 de mesure.

De plus, on s'intéresse ici à la structure induite par les formes musicales résultant des notes jouées et nous ne prenons pas en compte les informations sur les instants de silence. Ainsi, nous avons choisi de prolonger chaque note jouée jusqu'au début de la prochaine et donc d'étudier l'intervalle chromatique de ces notes sans tenir compte des silences insérés entre elles.

Ce traitement conduit certes à une perte d'information musicale qui pourrait s'avérer utile dans d'autres contextes d'étude, comme des analyses très fines du rythme ou des ornements mélodiques.

Dans le cadre de ces travaux, nous étudions essentiellement la forme et les relations multi-échelles présentes dans les mélodies au niveau de la partie basse des échelles mésoscopiques, et ces formes réduites sont considérées comme suffisantes à cet égard.

2.1.2 Surface musicale et segmentation

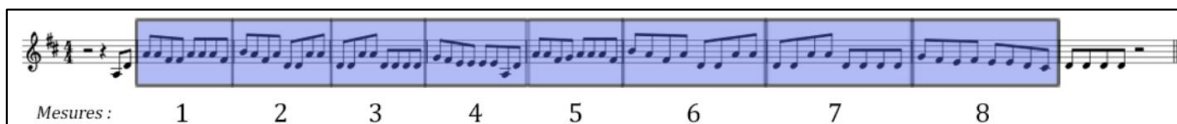


Figure 16: Portée d'une section musicale avec une surface de 8 mesures en version réduite.

Surface musicale

La ligne mélodique d'une section musicale telle que celle présentée sur la figure 16 ne se présente pas nécessairement de façon synchrone avec les barres de mesure de la portée.

En effet, la durée de la mélodie dans une section musicale n'est pas nécessairement synchronisée avec les informations métriques. Elle peut être soit anticipée, soit retardée par rapport à la barre de mesure initiale.

On définit la surface musicale comme la plage temporelle qui caractérise le début et la fin de la narration mélodique indépendamment de la grille métrique.

Ce phénomène est notamment observé lorsque la mélodie présente une anacrouse. La mélodie est ainsi désynchronisée par rapport aux frontières métriques, ainsi que par rapport à son harmonie sous-jacente.

Le phénomène d'anacrouse n'est pas très bien « balisé » dans la musicologie classique et encore moins pris en compte de façon précise en informatique musicale. Nous n'avons en effet pas trouvé dans la littérature de modèle satisfaisant qui rende compte de ce phénomène. Une contribution de ce travail consiste donc à aborder ce problème avec des outils issus de la Théorie de l'Information.

D'ailleurs, pour recomposer la mélodie, c'est la surface musicale qu'il convient de segmenter, afin de faire apparaître les frontières des éléments musicaux pertinents. En effet, plus les frontières des motifs définis par la segmentation sont en cohérence avec la surface musicale, plus on peut faire l'hypothèse que les motifs sont interchangeables, et qu'en conséquence les résultats des recompositions à partir de ces motifs seront pertinents.

Il existe également, mais plus rarement, des cas où la mélodie peut être retardée par rapport à barre de mesure et/ou faire se produire des débordements lorsqu'elle finit après la dernière barre de mesure.

Voici un exemple (Fig. 17) où la mélodie du morceau est désynchronisée par rapport à sa grille de mesure : dans cet exemple, la mélodie est anticipée selon un phénomène d'anacrouse qui se répète pour chaque mesure



Figure 17: Illustration du phénomène d'anacrouse et de débordement sur la partition de « Joyeux Anniversaire ».

Les barres diagonales, décalées par rapport aux barres de mesure, figurent les regroupements de notes en motifs au niveau de la surface musicale.

Afin de caractériser ces désynchronisations, on introduit une variable θ qui correspond au décalage de la surface musicale par rapport à la première barre de mesure exprimée en nombre de temps.

$$\theta \in [-2; +2] \text{ temps,}$$

avec un pas de 0,5 temps

Dans le présent travail, on considère que ce décalage est un décalage global et qui s'applique à l'ensemble de la section, bien que cette hypothèse soit clairement restrictive. Il existe en effet des cas où un décalage variable serait plus optimal, mais cela ajoute un degré de complexité supplémentaire qui n'est pas abordé dans ce travail.

Segmentation interne

Préalablement à la recomposition d'une section musicale, il convient de la segmenter en motifs.

Dans le cas le plus général, la tâche de segmentation d'un morceau de musique consiste à définir les motifs structurants d'une surface musicale à une certaine échelle (ou famille d'échelle).

Cette tâche revient à déterminer une grille de segmentation G permettant de découper la section en granules et couvrant l'ensemble de la section représentée à l'échelle des granules.

Cette grille correspond à un découpage de l'intervalle considéré en $z = 16$ motifs disjoints et contigus,

$$X = x_1 \dots x_t \dots x_z \quad 1 \leq t \leq z = 16 \quad (2.5)$$

Comme précisé ci-dessus, dans ces travaux, on se concentre uniquement sur les cas où la grille G est régulière, c'est-à-dire que chaque motif x_j est constitué exactement de 4 granules (quitte à laisser les granules surnuméraires en fin de section en dehors du périmètre de modélisation).

$$x_j = u_1^{(j)} \dots u_i^{(j)} \dots u_m^{(j)} \quad 1 \leq i \leq m = 4 \quad (2.6)$$

The figure illustrates the segmentation of a musical score into motifs and granules. The top part shows a musical score with 16 measures. The 10th measure is highlighted in red. Below the score, a red box labeled "Motif x_{10} " shows a zoomed-in view of the 10th measure, which contains four quarter notes. Below this box, the granules are labeled as $x_1^{(10)}$, $x_2^{(10)}$, $x_3^{(10)}$, and $x_4^{(10)}$.

Figure 18: Illustration des éléments granulaires constituant les motifs structurants. Ici, les 6 premières croches sont laissées hors du périmètre de modélisation du segment.

La grille régulière G peut éventuellement avoir un décalage à l'origine noté θ (défini dans la section précédente), on nomme cette grille G_θ .

Le processus de segmentation consiste donc à découper une section musicale X selon une grille de segmentation G_θ , permettant ainsi d'instancier $z = 16$ motifs qui varient avec le décalage θ .

Chaque motif décalé de θ s'écrit x_i^θ , et la section musicale notée X_θ peut s'écrire comme :

$$X_\theta = x_1^\theta \dots x_t^\theta \dots x_z^\theta \quad (2.7)$$

où x_t^θ est la version de x_t décalée de θ .

Chacun des motifs de cette segmentation est constitué de $m = 4$ granules, et on peut ainsi écrire X_θ de la façon suivante :

$$X_\theta = [u_{\theta+1} \dots u_{\theta+4}] \dots [u_{4d+\theta+1} \dots u_{4(d+1)+\theta}] \dots [u_{(g-3)+\theta} \dots u_{g+\theta}] \quad (2.8)$$

Cette étape de segmentation est essentielle pour étudier la structure d'une section musicale et a fortiori pour la mise en place de recompositions polytopiques.

En effet, elle conditionne la cohérence du réagencement des motifs de la section et de ce fait la musicalité potentielle du résultat.

Plusieurs procédés de segmentation sont considérés dans ce travail.

Segmentation manuelle : on peut tout d'abord faire appel à un expert pour réaliser une segmentation manuelle de la section X . De cette façon, les motifs structurants sont identifiés selon des considérations musicologiques plus ou moins formalisables, si tant est que l'expert les exploite correctement. Et sachant qu'il peut exister plusieurs solutions « défendables » entre lesquelles il est difficile d'arbitrer.

Segmentation algorithmique : une alternative, explorée dans cette thèse, consiste à développer un algorithme de segmentation dont le but est de déterminer un décalage θ^* , qui permet de trouver la valeur optimale de ce paramètre au sens d'un critère donné. Comme évoqué préalablement, ce critère sera, dans cette thèse, basée sur un coût de complexité.

Estimation des paramètres de segmentation

Déterminer les paramètres de segmentation est un problème d'estimation qui peut s'effectuer soit par une recherche itérative (par un processus d'optimisation), soit par une recherche exhaustive.

Dans ce dernier cas, on inventorie a priori toutes les segmentations candidates, en tant qu'ensemble $\{X_\theta\}$ comprenant tous les décalages possibles de la grille régulière par rapport à la première mesure et on évalue le score de chaque X_θ au sens d'un critère de coût retenu pour mesurer son « adéquation ».

Dans le cas présent, le décalage θ prend un nombre raisonnable de valeurs et le problème de l'optimisation algorithmique de θ peut donc être abordé par une recherche exhaustive.

Afin que la segmentation de la section musicale rende compte de possibles anacrouses ou de débordements, on autorise à ce que le début du premier motif commence en anticipation ou en retard par rapport à la première barre de mesure. On impose néanmoins un « décalage » maximal de $\frac{1}{2}$ mesure (= 2 temps) avant et après la première barre de mesure avec un pas de $\frac{1}{8}$ de mesure c'est-à-dire 1 granule.

$$\theta \in \{-2, -1.5, -1, -0.5, 0, 0.5, 1, 1.5, 2\} \text{ temps}$$

soient 9 valeurs possibles

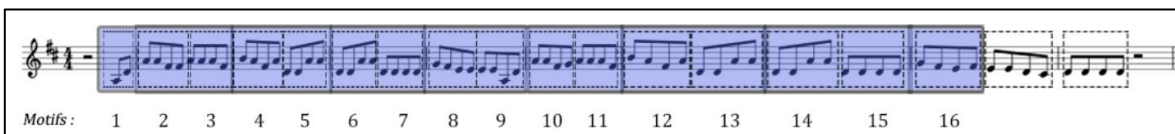


Figure 19: Segmentation régulière anticipée de 1/4 mesure ($\theta = -2$ granules = -1 temps)

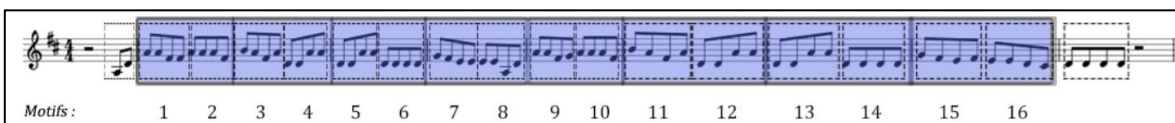


Figure 20. Segmentation régulière débutant sur la première mesure ($\theta = 0$ granule = 0 temps)

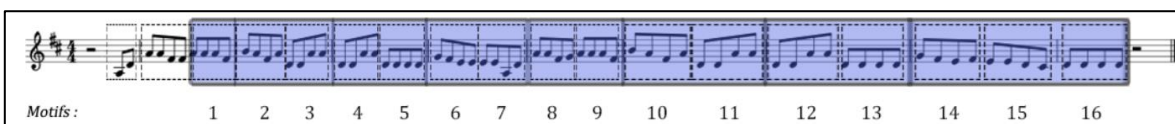


Figure 21: Segmentation régulière retardée de 1/2 mesure ($\theta = 4$ granules = 2 temps)

2.1.3 Graphe tesseractique

Cette partie illustre le passage de la segmentation d'une section X d'un point de vue séquentiel, au passage à une vue polytopique, sous forme de graphe de relations latentes.

Dans le cas présent, où la segmentation est régulière, la section mélodique considérée est de forme « carrée » sur 8 mesures.

Sur l'exemple qui suit, on suppose pour l'instant qu'elle n'est pas décalée par rapport aux informations métriques, soit un décalage $\theta = 0$ et cette section est donc nommée X_0 .

$$X_0 = (x_1^0, x_2^0, \dots, x_j^0, \dots, x_{16}^0) \quad (2.9)$$

On segmente ainsi une section musicale X en un ensemble de 16 motifs élémentaires structurants $\{x_1, x_2, \dots, x_{16}\}$ correspondant également aux 16 nœuds du graphe polytopique (ici, un tesseract) qui figure un ensemble de relations latentes.

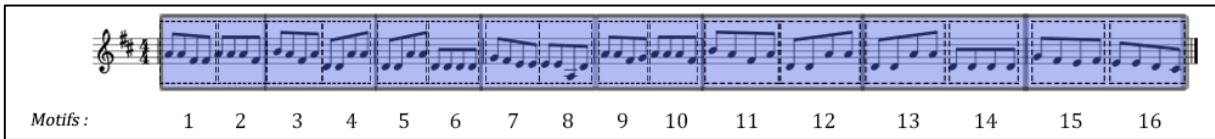


Figure 22: Segmentation X_0 en 16 motifs de la section X .

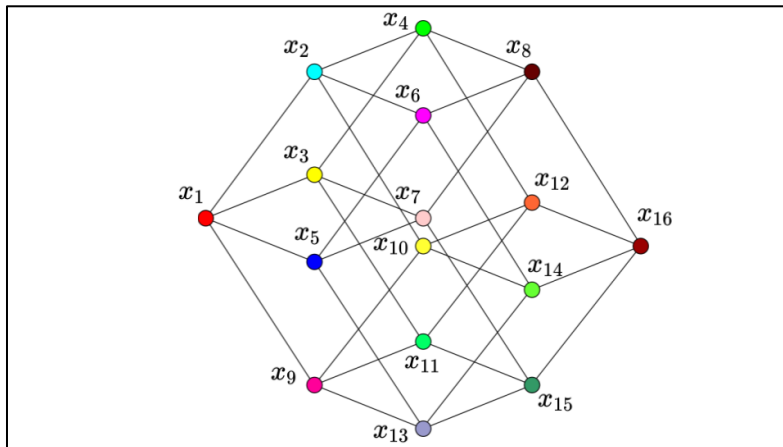


Figure 23: Visualisation des 16 motifs de la section sur son graphe polytopique (ici un tesseract)

Les arêtes du tesseract support mettent en évidence les relations entre les motifs en positions métriques homologues à plusieurs échelles temporelles.

Les sections suivantes détaillent l'estimation de ces relations multi-échelles et notamment lorsqu'on suppose qu'elles forment un système d'implication.

2.1.4 Systèmes d'implications au sein du modèle polytopique

Afin d'étudier les relations structurelles qu'entretiennent les motifs, nous allons nous intéresser à leur formalisation en termes de systèmes d'implication.

Au sein d'un système d'implication constitué de 4 motifs consécutifs, comme sur la figure ci-dessous, les relations entre motifs peuvent être décrites dans le cadre du modèle Système & Contraste.

- Le premier x_k initie le système et ne dépend d'aucun autre élément antécédent.
- Les sous-primier x_l et x_m présentent l'un, l'autre, les deux (ou aucun) des relations de similarité avec le premier et/ou entre eux.
- Le dernier élément x_n est souvent en disparité partielle avec les 3 éléments précédents, au sens où il se départit de la relation d'analogie attendue des 3 premiers éléments : il est appelé l'élément contraste du système.

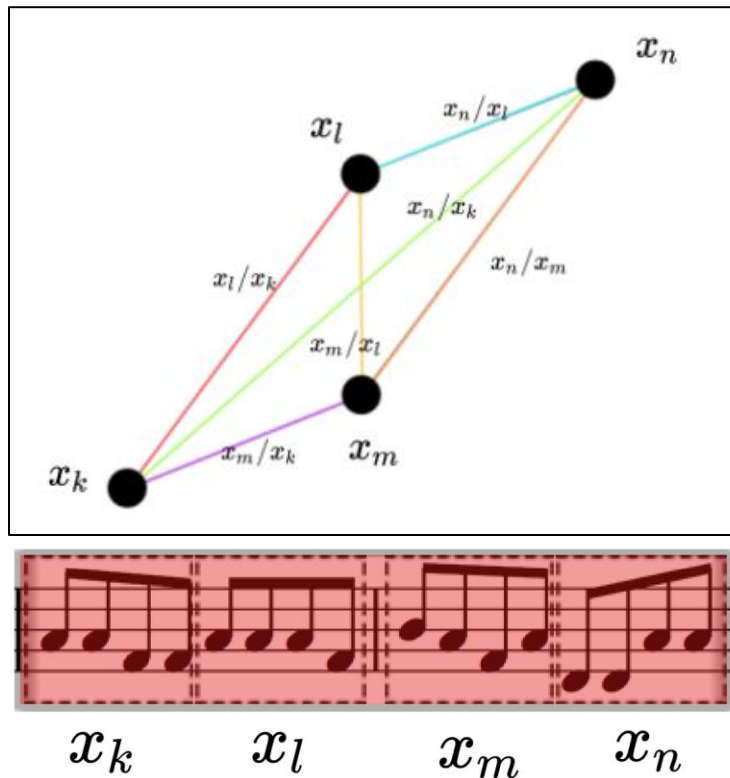


Figure 24: Illustration des relations présentes au sein d'un système

Différentes façons d'estimer les relations au sein d'un système peuvent être considérées et sont discutées en détail dans la section 3.1.3.

Selon nos hypothèses de travail, les motifs structurants d'une section musicale de 8 mesures sont organisés en sous-groupes de 4 motifs.

Ces ensembles de 4 motifs correspondent, dans la représentation tesseractique de la section X , à des groupes de sommets formant des faces parallèles.

Si l'on considère les 4 systèmes d'implications formés par les motifs apparaissant dans l'ordre séquentiel de la section X , on observe ces 4 systèmes sur les faces parallèles visualisées sur la figure 25.

Chacun des 4 groupes de motifs forme lui-même un système d'implication au sens du formalisme Système & Contraste (S&C) et les 4 groupes forment ainsi un système de système, on le désigne par la suite comme *Super-Système*.

La section musicale X est donc un ensemble de 4 systèmes d'implication, eux-mêmes constitués d'un ensemble de 4 motifs.

$$X = S_1 \dots S_k \dots S_q \quad 1 \leq k \leq q = 4 \quad (2.10)$$

$$S_k = x_1^{(k)} \dots x_j^{(k)} \dots x_n^{(k)} \quad 1 \leq j \leq n = 4 \quad (2.11)$$

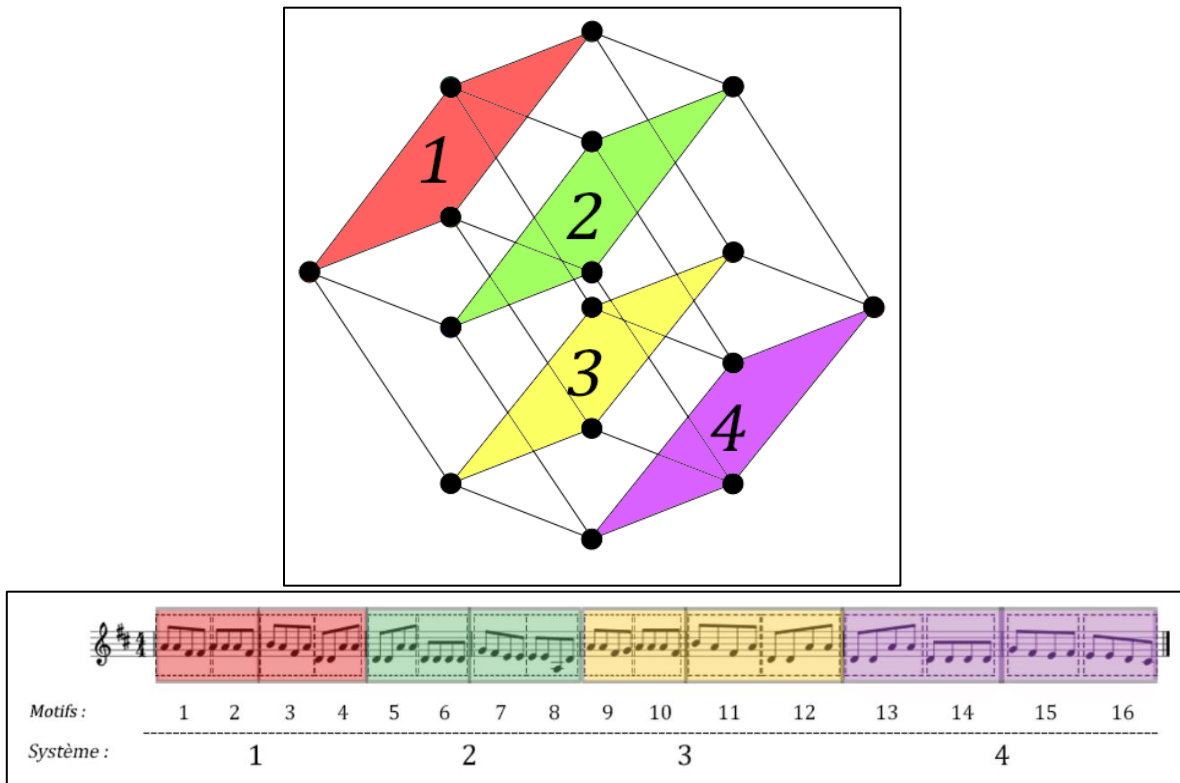


Figure 25: Représentation polytopique des 4 systèmes d'implications formés par les motifs dans l'ordre séquentiel.

En considérant les autres systèmes de faces parallèles, on remarque que la représentation polytopique fait apparaître d'autres décompositions possibles qui peuvent être utilisées de la même façon pour former 4 autres systèmes d'implications constitués de 4 motifs.

Les motifs regroupés selon ces nouveaux systèmes ne sont plus organisés selon l'ordre chronologique, mais sont constitués de motifs successifs non contigus en positions métriquement homologues, qui regroupent différents couples d'échelles de temps.

Dans le tesseract, on dénombre 6 systèmes distincts de faces parallèles à partir de ce support polytopique.

Ces différents systèmes de faces parallèles et les systèmes d'implication qui en résultent sont au cœur du concept de complexité polytopique pour les travaux présentés dans cette thèse.

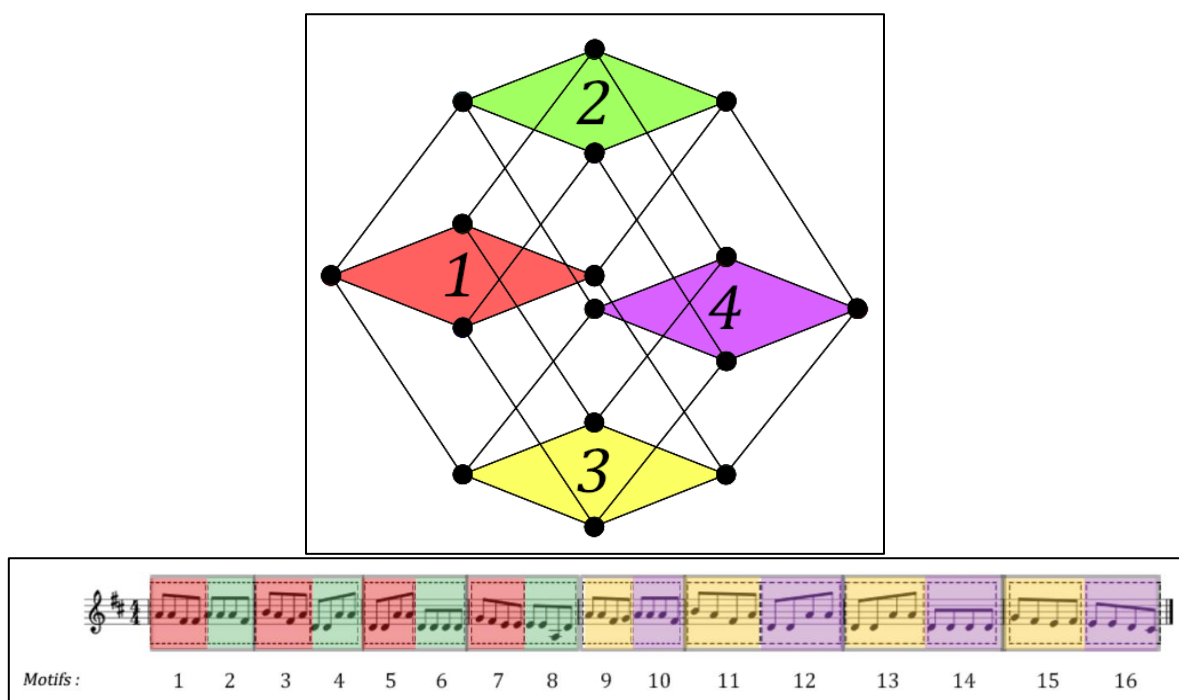


Figure 26: Représentation polytopique des 4 systèmes d'implications formés par les motifs selon un des systèmes de faces parallèles.

En effet, ces différentes configurations permettent la mise en correspondance de motifs distants, mais qui entretiennent des relations privilégiées du fait qu'ils sont homologues à certaines échelles temporelles.

De cette façon on pourra choisir d'estimer la complexité d'un motif à partir des motifs qui lui sont voisins dans l'ordre séquentiel, ou bien à partir des 5 autres systèmes de motifs homologues combinant d'autres échelles de temps, avec donc un caractère multi-échelle de l'optimisation du regroupement des motifs.

Chaque système de 4 faces parallèles du tesseract se définit par 2 directions d'arêtes à partir du premier motif de la segmentation (motif x_1) parmi $\lambda_1, \lambda_2, \lambda_3$ et λ_4 (représentées sur la Figure 27) ou $((\lambda_i, \lambda_j)_{i < j})$ définit de façon unique un ensemble de 4 faces parallèles formant les systèmes de bas-niveau dans le tesseract.

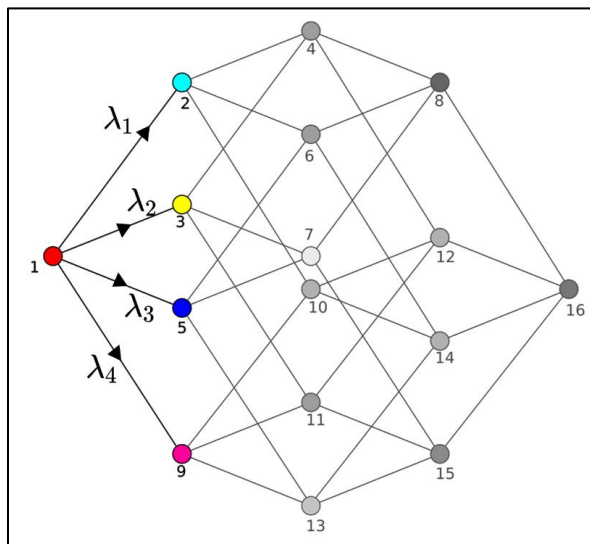


Figure 27 : Choix des directions permettant de caractériser les niveaux des systèmes d'implication dans le tesseract.

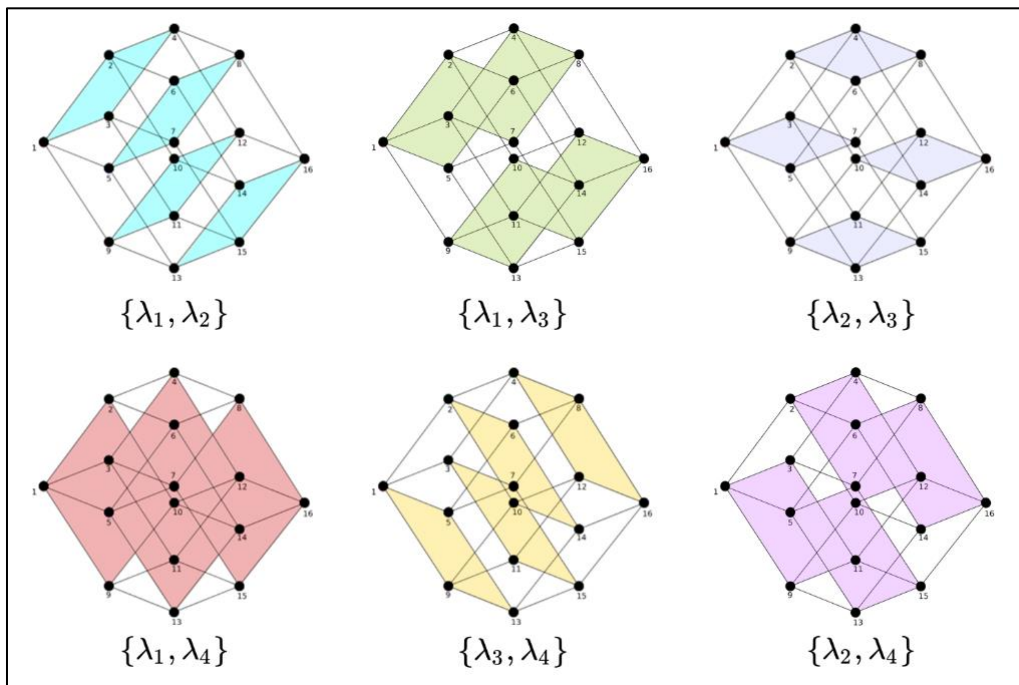


Figure 28: Représentation des 6 systèmes de faces parallèles du tesseract.

Pour un système de faces parallèles défini par $\{\lambda_i, \lambda_j\}$, 4 systèmes de bas-niveau sont donc déterminés, les directions d'arêtes restantes $\{\lambda_k, \lambda_\ell\}$ définissant une face formant un système à l'échelle supérieure contenant le prisme général x_1 . Ce sont ces systèmes que l'on dénomme les *Super-Systèmes*, et on en dénombre 6.

On a représenté en couleur sur chaque système de face parallèle, le 1^{er} super système reliant les éléments « primers » de chacun des 4 systèmes de bas-niveau pour former 1 seul système à l'échelle supérieure.

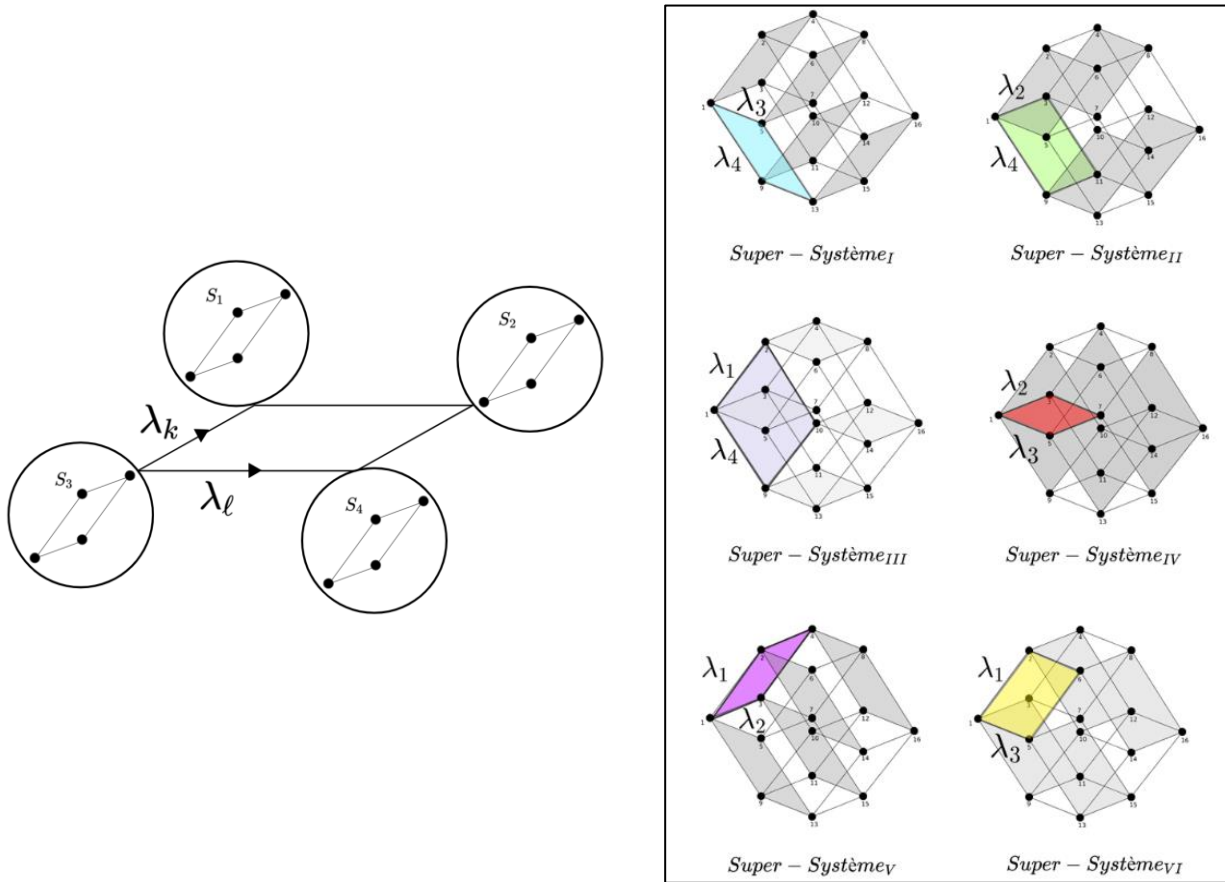


Figure 29: Illustration des Super Systèmes pour chacun des systèmes de faces parallèles.

Les différents systèmes de faces parallèles font apparaitre de nouvelles possibilités d'ordonnement de l'ordre temporel des motifs. Ces variations donnent lieu à un ensemble particulier de fonctions de permutations que l'on désigne par « Primer Preserving Permutations » ou PPP (Louboutin & Bimbot, 2017) et qui vont être également utilisés pour générer les recompositions polytopiques qui font l'objet de la section suivante.

2.2 Recompositions Polytopiques

2.2.1 Définition

Dans la section variations et recompositions, nous avons évoqué le principe des recompositions polytopiques en tant que variations de l'ordre temporel de motifs structurants. Nous définissons maintenant l'ensemble de fonctions qui permettent de réorganiser l'ordre des motifs d'une section musicale, dans le but d'en créer des variations conservant certaines propriétés de la structure originale, tout en recombinaison la hiérarchie des échelles.

Ainsi, les recompositions polytopiques telles que définies dans ce travail sont des permutations des indices temporels appliquées aux motifs structurants d'une section musicale qui respectent certaines propriétés.

Dans le cas général, une permutation est définie comme une fonction bijective φ appliquée sur les indices des z motifs (x_1, x_2, \dots, x_z) d'une section segmentée X (ayant éventuellement subi un décalage θ).

Une telle fonction de permutation φ est définie comme une bijection dans l'ensemble d'entiers $\{1, \dots, t, \dots, z\}$ et permet d'attribuer de nouvelles positions à chaque indice des motifs structurants de la section d'origine X .

$$\varphi = \{\varphi(1), \dots, \varphi(t), \dots, \varphi(z)\} \quad (2.12)$$

Cette fonction permet de définir le nouvel ordre dans lequel les motifs de la section d'origine \hat{X} vont être lus dans la section recomposée et on note alors :

$$X^\varphi = x_{\varphi(1)} \dots x_{\varphi(t)} \dots x_{\varphi(z)} \quad (2.13)$$

La nouvelle section obtenue X^φ est donc une nouvelle version, ou recomposition, de la section X d'origine où ses motifs sont agencés dans un ordre différent.

Ce processus peut également être expliqué comme le fait de choisir un chemin de parcours dans le polytope différent de celui induit par l'ordre séquentiel original. Ces fonctions ont donc pour effet de délinéariser la section musicale d'origine puis de la relinéariser avec un nouvel ordonnancement.

Les permutations engendrent néanmoins un certain nombre de « perturbations » dans le déroulement musical de la section d'origine et l'on discutera de la caractérisation qualitative et quantitative de ces perturbations dans la section 3.1.3.

2.2.2 Les PPP : des recompositions particulières

Il existe précisément 16! (soit un peu plus de 20 000 milliards de) permutations φ possibles pour réorganiser 16 motifs dans un ordre différent de l'ordre d'origine.

Dans cette section, nous définissons un (tout petit) sous-ensemble de permutations permettant de préserver les propriétés caractéristiques des structures formulées dans les hypothèses de travail et en particulier d'assurer l'invariance des coûts de complexité des relations induites par les systèmes d'implication.

On obtient ce sous-ensemble en caractérisant les permutations qui intervertissent les échelles des relations entre motifs métriquement homologues et ainsi qui interchangent les faces du tesseract.

On retrouve ainsi les PPP qui consistent à choisir deux directions d'arêtes λ_i et λ_j sur le polytope parmi $\lambda_1, \lambda_2, \lambda_3$ et λ_4 dont la seconde est d'indice plus grand que la première ($j > i$) soit 6 choix possibles.

$$\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\} \rightarrow \{\lambda_i, \lambda_j, \lambda_k, \lambda_\ell\} \text{ avec :}$$

$$i \neq j \neq k \neq \ell \text{ (du fait qu'il s'agit d'une permutation)}$$

$$i < j \text{ et } k < \ell \text{ (préservation d'un ordre relatif des échelles)}$$

Les 2 directions λ_i et λ_j définissent les faces parallèles pour les 4 systèmes de bas niveau et la transformation qui amène le système de faces initiales sur ce nouveau système de faces en préservant l'ordre temporel naturel dans chaque face est notée π_p avec $0 \leq p \leq 5$

En appliquant la permutation π_p à la séquence X , on obtient:

$$X^{\pi_p} = x_{\pi_p(1)}, x_{\pi_p(2)}, \dots, x_{\pi_p(16)} \quad 0 \leq p \leq 5 \quad (2.14)$$

que l'on notera également $x_1[p], x_2[p], \dots, x_{16}[p]$

La figure 30 donne les indices des permutations π_p

	$\pi_p(1)$	$\pi_p(2)$	$\pi_p(3)$	$\pi_p(4)$	$\pi_p(5)$	$\pi_p(6)$	$\pi_p(7)$	$\pi_p(8)$	$\pi_p(9)$	$\pi_p(10)$	$\pi_p(11)$	$\pi_p(12)$	$\pi_p(13)$	$\pi_p(14)$	$\pi_p(15)$	$\pi_p(16)$
π_0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
π_1	1	3	5	7	2	4	6	8	9	11	13	15	10	12	14	16
π_2	1	2	9	10	3	4	11	12	5	6	13	14	7	8	15	16
π_3	1	2	5	6	3	4	7	8	9	10	13	14	11	12	15	16
π_4	1	5	9	13	2	6	10	14	3	7	11	15	4	8	12	16
π_5	1	3	9	11	2	4	10	12	5	7	13	15	6	8	14	16

Figure 30: Définition des permutations (polytopiques) préservant les primers (PPP).

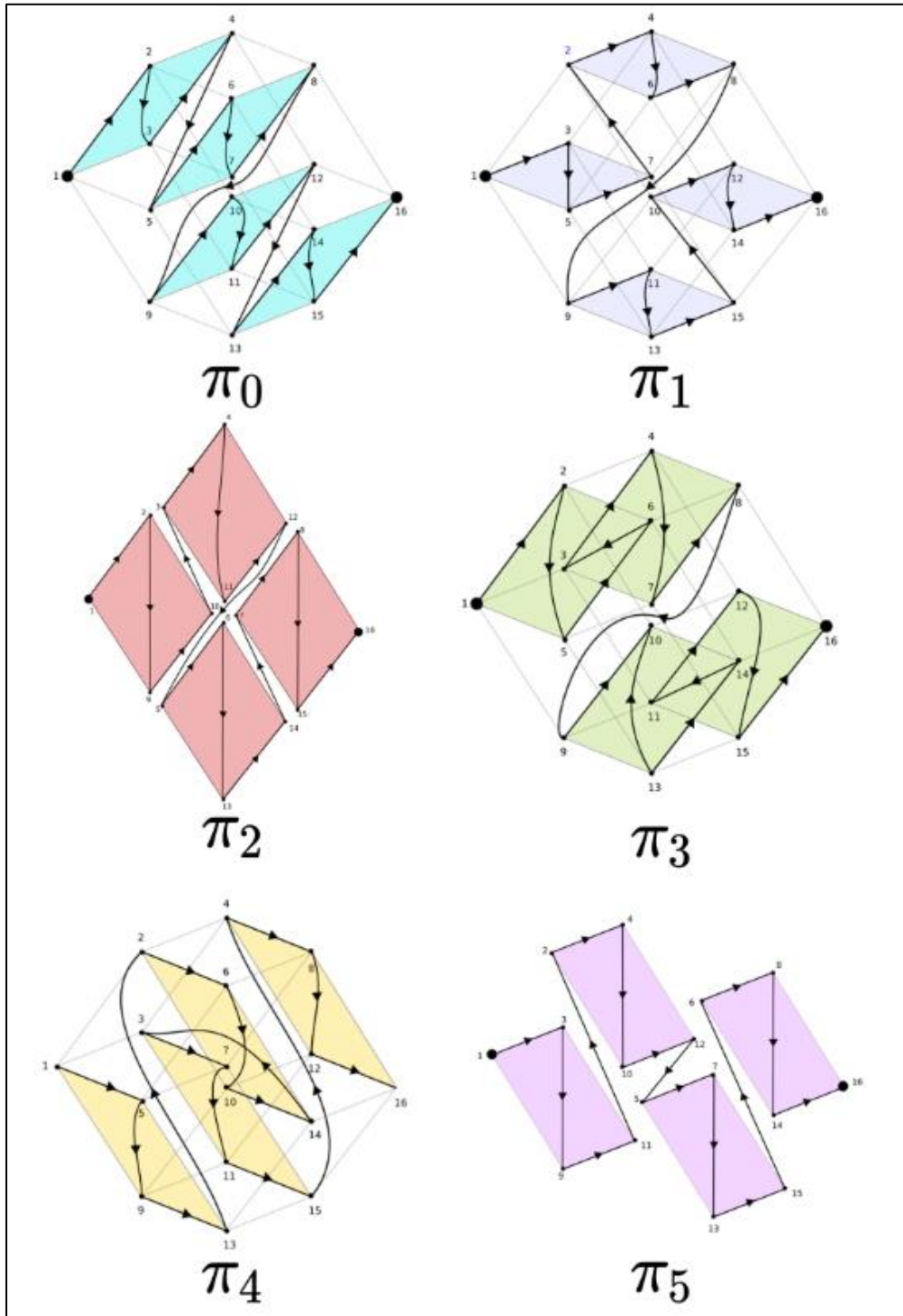


Figure 31: Illustration des parcours du tesseract selon les différentes PPP.

L'exemple de la figure 32 vient illustrer comment les PPP préservent les relations systémiques entre motifs musicaux dans la section en interchangeant les échelles temporelles auxquelles elles se développent.

Ainsi, les demi-mesures qui se situent en position 1, 3, 9 et 11 dans l'original se retrouvent voisines dans la 6ème PPP, π_5 .

L'utilisation des PPP comme permutations privilégiées est destinée à préserver la chronologie des motifs musicaux dans les systèmes de bas niveau bien qu'ils soient issus d'échelles temporelles différentes dans l'original.

Ces recompositions induisent un degré variable de perturbation qui est discuté dans la sous-section qui suit.

The figure displays six musical staves, each representing a different permutation (π_0 through π_5) of a melody. The melody is written in a 4/4 time signature with a key signature of one sharp (F#). The original melody (π_0) consists of the following notes: D4, F#4, A4, B4, C5, B4, A4, G4, F#4, E4, D4, C4. The other staves (π_1 to π_5) show various rearrangements of these notes, illustrating how the polytopical permutations preserve the systemic relationships between motifs while changing their temporal positions.

Figure 32: Exemple des 5 recompositions polytopiques PPP appliquées à la mélodie d'un extrait musical de la base RWC POP.

2.2.3 Caractérisation des recompositions

Afin de caractériser l'impact que provoquent les recompositions polytopiques sur les propriétés structurelles des versions recomposées, nous allons préciser et quantifier ces perturbations d'un point de vue topologique.

Par topologique, nous entendons que cette caractérisation s'effectue indépendamment du contenu musical effectif réorganisé par les fonctions de recompositions, dont les similarités particulières pourraient influencer l'impact.

Les perturbations d'ordre topologiques inhérentes à chacune des PPP peuvent résulter de deux caractéristiques de ces permutations

- D , le nombre de discontinuités par rapport à la séquence d'origine résultant de la recomposition, et
- E , l'excursion d'échelle, c'est-à-dire le (\log_2) de l'intervalle temporel maximum entre deux indices d'origine d'éléments consécutifs dans la séquence recomposée.

$$D = \# \{ t \mid [\pi(t) - \pi(t - 1)] \neq 1 \} \quad (0 < t < N) \quad (2.15)$$

$$E = \log_2 \max_{0 < t < N} |\pi(t) - \pi(t - 1)| \quad (2.16)$$

Les recompositions PPP induisent entre 7 et 16 discontinuités de linéarité entre la section d'origine et la version recomposée (voir tableau fig. 35). La figure 33 illustre les discontinuités résultant de la recomposition par π_2 .

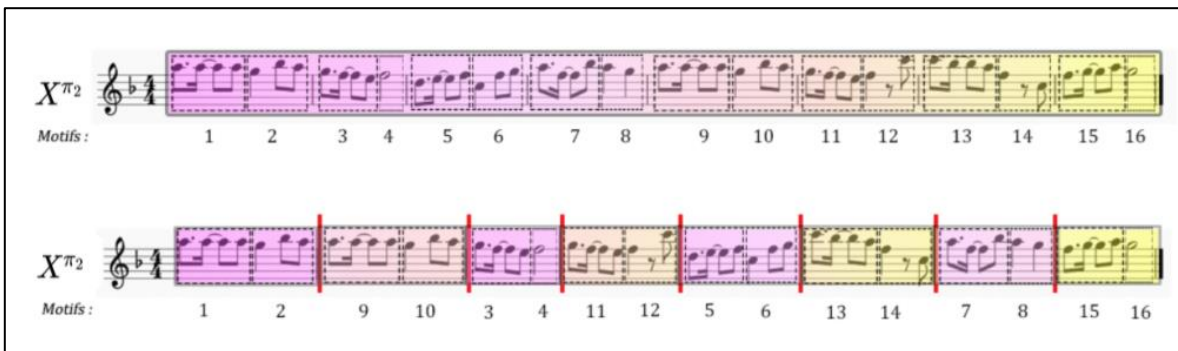


Figure 33: Illustration des 7 discontinuités induites par la recomposition π_2 .

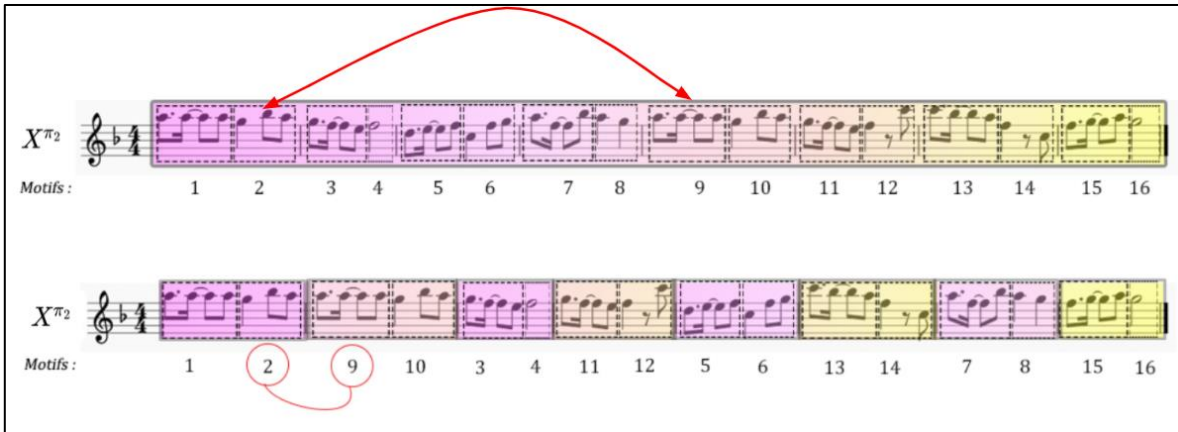


Figure 34: Illustration de la caractéristique d'excursion maximale $E = 3$, induite par la fonction de recomposition π_2 .

L'excursion, pour sa part, varie entre 2 et 4 selon les fonctions de recomposition PPP. Le tableau de la Figure 35 résume les valeurs de discontinuité (Disc.) et d'excursion (Exc.) pour les 6 PPP.

	$\pi_p(1)$	$\pi_p(2)$	$\pi_p(3)$	$\pi_p(4)$	$\pi_p(5)$	$\pi_p(6)$	$\pi_p(7)$	$\pi_p(8)$	$\pi_p(9)$	$\pi_p(10)$	$\pi_p(11)$	$\pi_p(12)$	$\pi_p(13)$	$\pi_p(14)$	$\pi_p(15)$	$\pi_p(16)$	Disc.	Exc.
π_0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1	2
π_1	1	3	5	7	2	4	6	8	9	11	13	15	10	12	14	16	15	2
π_2	1	2	9	10	3	4	11	12	5	6	13	14	7	8	15	16	7	3
π_3	1	2	5	6	3	4	7	8	9	10	13	14	11	12	15	16	7	2
π_4	1	5	9	13	2	6	10	14	3	7	11	15	4	8	12	16	16	4
π_5	1	3	9	11	2	4	10	12	5	7	13	15	6	8	14	16	16	3

Figure 35: Caractéristiques topologiques des 6 PPP.

Dans la suite de ce travail, nous allons mesurer, par le biais de tests perceptifs, l'impact de ces différentes PPP, en lien avec les perturbations topologiques qu'elles induisent, et par comparaison avec d'autres permutations qui ne respectent pas les invariances des PPP.

2.3 Conclusion

Dans ce chapitre, nous avons défini les différentes échelles temporelles et les objets musicaux étudiés dans ces travaux.

Il a été montré qu'une étape essentielle de cette étude était d'instancier de manière optimale les motifs structurants d'une section musicale.

Il est possible de réaliser cette étape de segmentation manuellement, mais un des objectifs de ces travaux est de rendre ce processus automatique par optimisation d'un critère de complexité et qui sera discuté dans le chapitre suivant.

Une fois les motifs structurants instanciés par segmentation manuelle ou automatique, on met en œuvre des scénarios de variations temporelles de ces motifs, qui sont dénommées recompositions polytopiques.

Pour cela, 6 fonctions bijectives particulières dites PPP ont été définies.

Elles conservent la logique induite par la structure polytopique dans sa forme hypercubique tout en intervertissant les échelles temporelles des relations entre motifs en positions métriques homologues. Elles ont finalement pour effet de réarranger les motifs structurants de la section d'origine dans un ordre temporel différent.

Ces fonctions induisent un certain nombre de « perturbations » et dont l'analyse a été présentée à la fin de ce chapitre.

Le chapitre suivant présente le critère de complexité utilisé pour la segmentation automatique ainsi que quelques détails spécifiques à la mise en œuvre algorithmique des processus qui viennent d'être présentés.

Chapitre 3

Critère de complexité et algorithme de segmentation

Ce chapitre détaille l'utilisation de critères de complexité utilisés pour établir automatiquement la segmentation d'un extrait musical, certains aspects de la mise en œuvre algorithmique de ces concepts et fournit une première évaluation des méthodes considérées. On rappelle que, dans ces travaux, on étudie la structure de la ligne mélodique d'un refrain de musique pop et les relations que ses motifs structurants entretiennent à plusieurs échelles temporelles. Ces relations sont représentées à l'aide du support polytopique de la section musicale sous la forme d'un 4-cube (ou tesseract).

Ce support correspond au graphe de relations latentes explicité dans les travaux de Louboutin (Louboutin, 2019) et une des tâches essentielles des travaux de la présente thèse est d'instancier de façon optimale les nœuds de ce graphe par des motifs mélodiques en tenant compte du fait que la surface musicale de ces motifs n'est pas nécessairement synchronisée avec les barres de mesure.

La détermination de ces motifs s'effectue par un processus de segmentation de la section musicale au sens d'un critère de complexité multi-échelle qui repose sur l'estimation d'une complexité « algorithmique » au sens de Kolmogorov et qui estime la charge de traitement nécessaire à l'explication d'une section musicale sous différentes hypothèses détaillées dans ce chapitre.

Après avoir présenté ces différentes variantes d'hypothèses, nous présentons le processus global de détermination d'une segmentation « optimale », c'est-à-dire la façon dont s'articulent les critères définis aux différentes échelles, et la combinatoire qui en résulte.

Nous détaillons ensuite l'implémentation de l'algorithme utilisé pour la génération des recompositions polytopiques d'une section musicale.

Finalement, on décrit une méthodologie adoptée de façon préliminaire pour évaluer les méthodes d'estimation algorithmique en les comparant à deux jeux d'annotations manuelles.

3.1 Critère de complexité

Dans la partie précédente, on a montré qu'il est possible d'utiliser plusieurs décalages θ de la grille régulière G par rapport à la première barre de mesure pour segmenter une même séquence musicale X avec différents décalages.

Dans cette section, nous allons préciser des critères pour caractériser une telle segmentation X_θ par un critère estimant la complexité de sa description, ces critères étant utilisés ensuite comme critères d'optimisation lors de la mise en place du processus de segmentation.

Le but de ce critère de complexité est donc de définir le décalage qui délimite le mieux les motifs structurants d'une section musicale.

Selon les hypothèses de ces travaux, le caractère structurant d'un motif musical x_t est évalué par son degré de similarité avec les motifs de la séquence déjà survenus. Autrement dit, plus le motif x_t présente des similarités avec des motifs antérieurs en position métrique homologue et/ou proche, plus il est considéré comme jouant un rôle structurant dans la narration musicale.

Plusieurs méthodes d'estimations (dites « *configurations* ») sont envisagées pour calculer cette complexité, selon différentes options qui sont inventoriées ci-dessous et détaillées ensuite :

- (1) La complexité de l'ensemble des 4 systèmes d'implication peut s'estimer en considérant chacun des 4 systèmes comme indépendants ou en prenant en compte l'influence du système à l'ordre supérieur reliant ces 4 systèmes (via leurs 4 primers).
- (2) On pose ensuite différentes hypothèses quant au choix de l'historique pertinent $H(x_t)$ d'un motif x_t au sein d'un système.
- (3) Plusieurs hypothèses sont également possibles par rapport à la dépendance des granules au sein d'un motif, ce qui impacte la façon de propager la ou les différences entre granules homologues au reste du motif.
- (4) On considère différentes façons de définir un granule de référence dans les motifs, c'est-à-dire celui à partir duquel on ancre le calcul d'une mesure de distorsion entre motifs. Si le granule de référence est le 1^{er} granule du motif, le calcul se fera de gauche à droite, mais si c'est un autre granule, le calcul de la distorsion s'opèrera de façon bilatérale.
- (5) Enfin, la distance élémentaire entre deux granules peut être définie de plusieurs façons. En particulier, par l'intervalle en nombre de demi-tons entre les notes des granules où par une distance favorisant des intervalles de notes ayant des harmoniques communs.

3.1.1 Complexité d'une section musicale

L'évaluation du degré de redondance présent au sein d'une section musicale segmentée X peut s'aborder comme l'estimation de sa complexité $C(X)$ dans un paradigme de « type complexité algorithmique ».

En toute généralité, la complexité d'une séquence $X = x_1 \dots x_t \dots x_z$ peut se décomposer en coûts conditionnels des motifs successifs :

$$\begin{aligned} C(X) &= C(x_1, \dots, x_t, \dots, x_z) \\ &= C(x_1) + C(x_2|x_1) + C(x_3|x_1x_2) + \dots + C(x_t|x_1 \dots x_{t-1}) + \dots + C(x_z|x_1 \dots x_{z-1}) \end{aligned} \quad (3.1)$$

Dans cette expression, on considère que la complexité d'un nouvel élément peut se calculer conditionnellement à tous les éléments antécédents.

Cependant, les complexités conditionnelles avec un grand horizon temporel sont difficiles à évaluer et on simplifie le calcul en approximant la complexité d'un nouveau motif x_t en se limitant à un sous-ensemble restreint d'éléments antécédents que l'on appelle son historique $H(x_t)$.

Nous approximons l'expression de la complexité de la façon suivante :

$$C(X) \approx \sum_{1 \leq t \leq z} C(x_t|H(x_t)) \quad (3.2)$$

où $H(x_t)$ désigne l'historique du motif x_t

Les différentes façons de définir H sont autant d'hypothèses sur le modèle structurel de la section selon les hypothèses de dépendances qu'entretiennent les motifs entre eux.

Dans le cadre du modèle polytopique, on fait l'hypothèse de grouper les motifs structurants en systèmes d'implication de 4 motifs pour définir des historiques qui simplifieront le calcul de la complexité.

On définit donc tout d'abord la complexité d'une section en fonction de sa décomposition en systèmes puis la complexité d'un système proprement dit.

3.1.2 Estimation de la complexité entre les systèmes

Selon l'approche adoptée dans cette thèse, une section musicale segmentée X peut-être caractérisée comme un ensemble de 4 systèmes choisis parmi un ensemble de systèmes de faces parallèles définis par deux directions d'arêtes dans $\{\lambda_1, \lambda_2, \lambda_3 \text{ et } \lambda_4\}$ à partir du motif x_1 .

$$X = S_1 \dots S_k \dots S_q \quad 1 \leq k \leq q = 4 \quad (3.3)$$

La complexité de cette séquence est donc relative à la complexité de ses systèmes constituants.

$$C(X) = C(S_1, S_2, S_3, S_4) \quad (3.4)$$

Plusieurs hypothèses sont envisageables pour calculer cette complexité.

On considère ici deux hypothèses d'estimation différentes : l'hypothèse selon laquelle chaque système de la section X est indépendant et l'hypothèse où le système d'ordre supérieur induit des dépendances supplémentaires.

Hypothèse d'indépendance des systèmes

La première hypothèse que l'on considère est que les 4 systèmes de bas niveau sont indépendants dans le calcul de la complexité, soit :

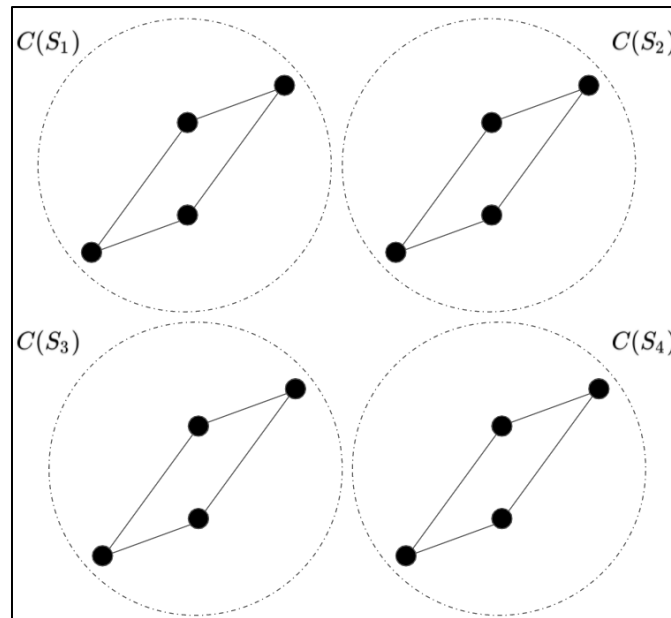


Figure 36: Illustration du calcul de complexité lorsque les 4 systèmes d'implications sont considérés indépendants.

$$\begin{aligned} \mathcal{C}(X) &= \mathcal{C}(S_1, S_2, S_3, S_4) \\ &\approx \mathcal{C}(S_1) + \mathcal{C}(S_2) + \mathcal{C}(S_3) + \mathcal{C}(S_4) \end{aligned} \quad (3.5)$$

$$= \mathcal{C}\left(x_1^{(1)} x_2^{(1)} x_3^{(1)} x_4^{(1)}\right) + \mathcal{C}\left(x_1^{(2)} x_2^{(2)} x_3^{(2)} x_4^{(2)}\right) + \mathcal{C}\left(x_1^{(3)} x_2^{(3)} x_3^{(3)} x_4^{(3)}\right) + \mathcal{C}\left(x_1^{(4)} x_2^{(4)} x_3^{(4)} x_4^{(4)}\right)$$

On note qu'il s'agit d'un cas particulier de la formule (3.2) où $H\left(x_j^{(k)}\right) = \left\{x_h^{(k)}\right\}_{1 \leq h < j}$

Hypothèse d'indépendance des systèmes + système à l'ordre supérieur

Dans ce second cas de figure, en plus du calcul de complexité de chacun des systèmes d'implication, on ajoute la complexité liée au super-système d'ordre supérieur formé par les « primers » des 4 systèmes de haut niveau que l'on note SS .

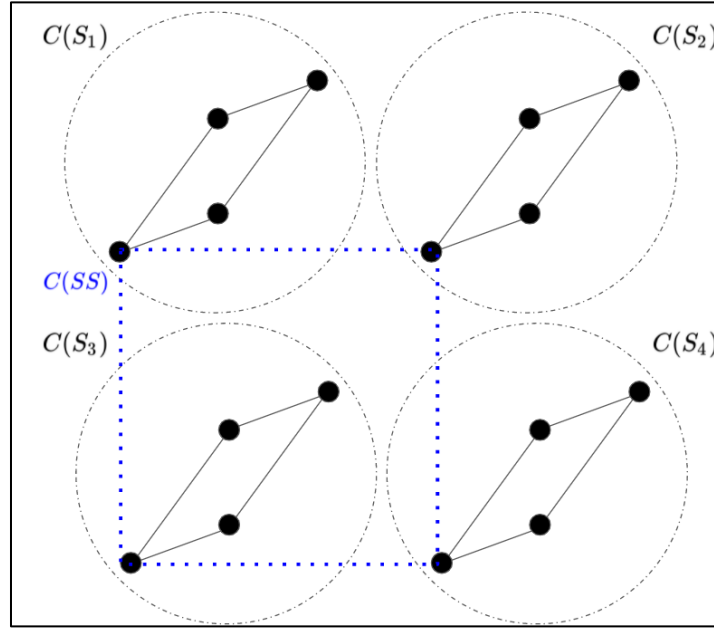


Figure 37: Illustration du calcul de complexité lorsque l'on prend en compte la complexité liée au Super-Système.

$$\begin{aligned} \mathcal{C}(X) &\approx \mathcal{C}(S_1) + \mathcal{C}(S_2) + \mathcal{C}(S_3) + \mathcal{C}(S_4) + \mathcal{C}(SS) \\ &= \mathcal{C}\left(x_1^{(1)} x_2^{(1)} x_3^{(1)} x_4^{(1)}\right) + \mathcal{C}\left(x_1^{(2)} x_2^{(2)} x_3^{(2)} x_4^{(2)}\right) + \mathcal{C}\left(x_1^{(3)} x_2^{(3)} x_3^{(3)} x_4^{(3)}\right) + \mathcal{C}\left(x_1^{(4)} x_2^{(4)} x_3^{(4)} x_4^{(4)}\right) \\ &\quad + \mathcal{C}\left(x_1^{(1)} x_1^{(2)} x_1^{(3)} x_1^{(4)}\right) \end{aligned} \quad (3.6)$$

3.1.3 Complexité d'un système

Après avoir formulé les hypothèses quant à l'estimation de la complexité selon l'organisation et les dépendances entre les systèmes d'implications, nous détaillons maintenant comment est définie la complexité d'un système donné.

Un système est constitué d'une séquence de quatre motifs en positions métriques homologues se produisant dans un ordre temporel croissant mais pas nécessairement constitué d'éléments consécutifs.

$$S = x_k x_l x_m x_n \text{ avec } k < l < m < n \quad (3.7)$$

La complexité conditionnelle de chaque motif est évaluée par rapport à son historique H , constituée d'un ou plusieurs motifs antécédents et il existe différentes possibilités pour définir cet historique.

Dans ce chapitre, nous allons détailler 3 hypothèses considérées dans cette thèse.

Estimation de la complexité d'un système selon toutes ses dépendances

Une façon d'estimer la complexité d'un système est de l'estimer comme étant une somme pondérée des complexités conditionnelles de tous les couples de motifs en respectant leur ordre causal :

$$\mathcal{C}(S) \approx \alpha_1 + \alpha_2 \mathcal{C}(x_l | x_k) + \alpha_3 (\mathcal{C}(x_m | x_k) + \mathcal{C}(x_m | x_l)) + \alpha_4 (\mathcal{C}(x_n | x_k) + \mathcal{C}(x_n | x_l) + \mathcal{C}(x_n | x_m)) \quad (3.8)$$

où les variables α_i correspondent à la pondération associée à chacun des motifs du système. On précise que α_1 est un terme constant qui n'impacte pas la détermination du coût minimum

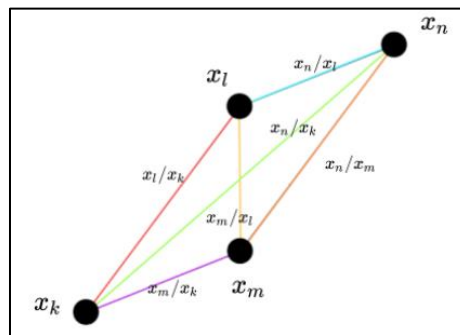


Figure 38: Relations de dépendances au sein d'un système d'implication pour l'estimation de sa complexité.

On tient compte ici de toutes les relations possibles à l'intérieur d'un système à 4 éléments.

Estimation de la complexité par rapport au premier élément du système :

Dans ce second cas de figure, on considère que le premier élément du système (primer) est le seul élément antécédent des autres éléments du système, soit :

$$\mathcal{C}(S) \approx \mathcal{C}(x_l|x_k) + \mathcal{C}(x_m|x_k) + \mathcal{C}(x_n|x_k) \quad (3.9)$$

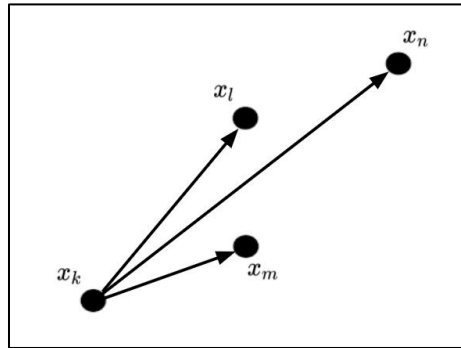


Figure 39: Illustration de l'estimation d'un système d'implication par rapport au premier élément.

Estimation par sélection du meilleur antécédent de l'élément contraste

Dans cette 3^{ème} méthode, on considère, comme dans la méthode précédente, que le primer correspond à l'historique des éléments 2 et 3.

Par contre, du fait que l'élément contraste x_n a un comportement supposé être différent des 3 premiers éléments, on détermine l'élément antécédent comme étant celui qui est le plus similaire parmi les trois précédents.

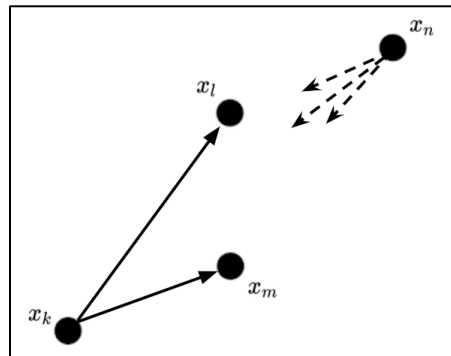


Figure 40: Illustration de l'estimation d'un système d'implication avec recherche du meilleur antécédent de l'élément contraste.

$$\mathcal{C}(S) \approx \mathcal{C}(x_l|x_k) + \mathcal{C}(x_m|x_k) + \min\{\mathcal{C}(x_n|x_k), \mathcal{C}(x_n|x_l), \mathcal{C}(x_n|x_m)\} \quad (3.10)$$

3.1.4 Complexité conditionnelle d'un motif

Chaque terme du calcul de la complexité entre systèmes est une valeur de complexité conditionnelle $\mathcal{C}(y|x)$ d'un motif y , selon son antécédent x , mesurant la distorsion (ou le surcroît d'information) nécessaire pour expliquer y connaissant x .

Dans le présent travail, cette complexité se décompose selon un calcul de transport entre les granules constituant chacun des deux motifs. Cela revient à considérer une famille de programme dont chaque instruction est un transport « élémentaire » de l'énergie d'un granule source vers un granule cible.

On considère deux motifs x et y chacun constitué respectivement de m et n granules.

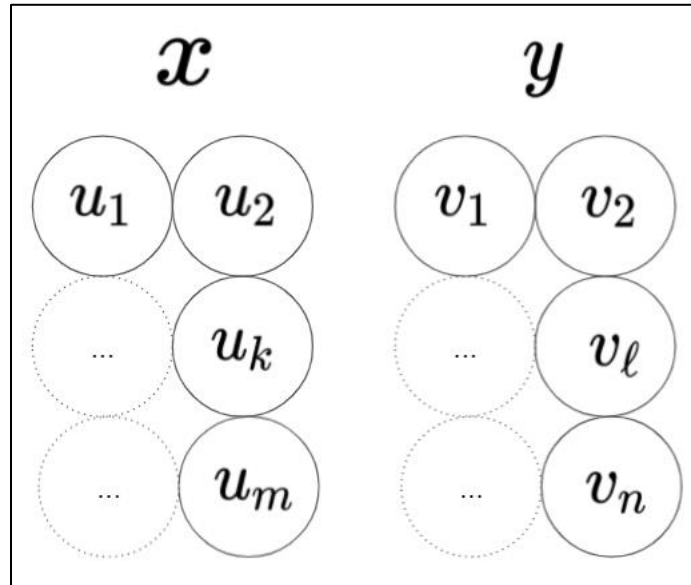


Figure 41: Représentation des motifs x et y constitué respectivement de m et n granules

L'expression de la complexité conditionnelle est la suivante :

$$\mathcal{C}(y|x) = \mathcal{C}(v_1 \dots v_l \dots v_n | u_1 \dots u_k \dots u_m) \quad (3.11)$$

Dans les travaux de cette thèse, nous considérons que les deux motifs x et y sont de même taille et constitués de 4 granules, ce qui correspond à la taille typique des motifs de notre étude (soit, $\frac{1}{2}$ mesure).

$$n = m = 4$$

$$x = u_1 u_2 u_3 u_4 \text{ et } y = v_1 v_2 v_3 v_4 \quad (3.12)$$

$$\mathcal{C}(y|x) = \mathcal{C}(v_1 v_2 v_3 v_4 | u_1 u_2 u_3 u_4)$$

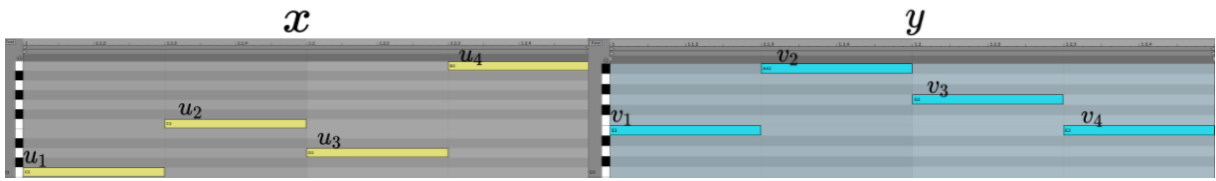


Figure 42: Exemple de deux motifs x et y dans leur représentation Piano Roll.

À l'échelle de l'estimation de la complexité d'un motif par rapport à son antécédent, 2 stratégies d'estimations sont étudiés : la façon de propager la dépendance entre granules et la définition du granule d'ancrage pour la recherche. Ces variantes sont présentées dans les sections suivantes.

3.1.4.1 Modèle de propagation des dépendances

(a) Indépendance des granules

Une première hypothèse pour calculer la complexité d'un motif par rapport à son antécédent est de calculer indépendamment la complexité de chaque granule du motif.

$$\mathcal{C}(y|x) \approx \mathcal{C}(v_1|u_1) + \mathcal{C}(v_2|u_2) + \mathcal{C}(v_3|u_3) + \mathcal{C}(v_4|u_4) \quad (3.13)$$

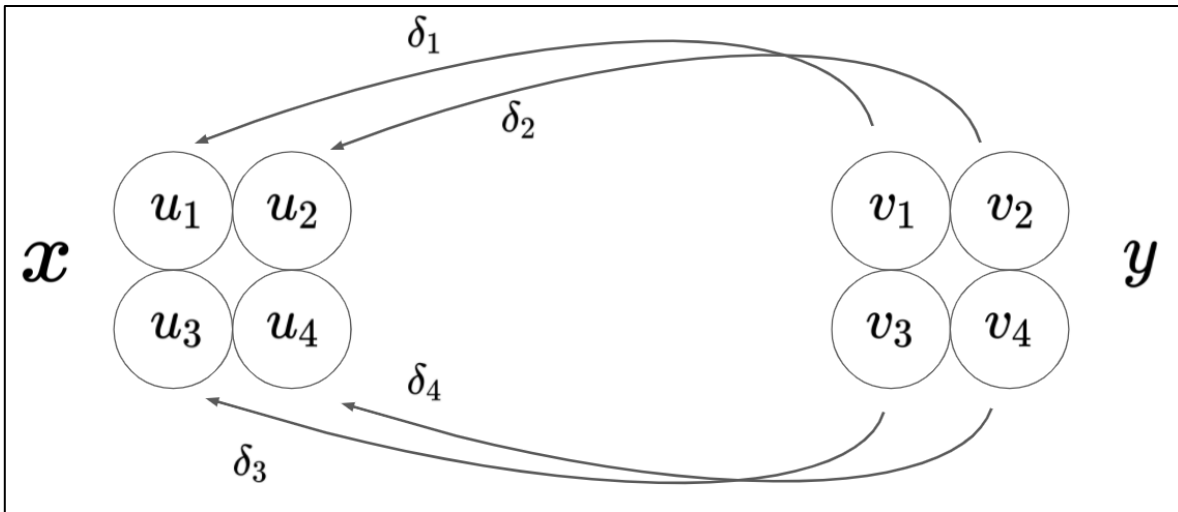


Figure 43: Expression de la complexité $\mathcal{C}(y|x)$ sous l'hypothèse de l'indépendance des granules constituants.

Chaque terme élémentaire dans le calcul de la complexité conditionnelle globale s'obtient donc comme la complexité d'un granule par rapport au granule en position homologue dans le motif antécédent.

Dans ce travail, cette quantité élémentaire estimée à partir de la différence de hauteur entre les 2 notes constituant de chacun de ces granules et est exprimée soit en nombre de $\frac{1}{2}$ tons (variante 1) soit à partir d'une distance proportionnelle au transport d'énergie entre représentations schématiques des harmoniques entre les 2 notes (variante 2).

Dans les deux cas, on appelle δ_i , l'intervalle chromatique.

La complexité entre deux motifs dépend donc de cet intervalle et est exprimée comme suit :

$$\delta_i = u_i - v_i$$

$$C(v_i|u_i) = C(v_i - u_i) = C(\delta_i) \quad (3.14)$$

où δ_i est l'intervalle chromatique exprimé en demi-ton entre les notes constituant des deux granules.

(b) Propagation de l'intervalle

L'hypothèse d'indépendance des granules est une hypothèse forte qui présente un certain nombre de limitations.

Par exemple, dans le cas où les deux motifs sont identiques à une translation fréquentielle près, la méthode précédente ne tient pas compte la forte similarité entre les motifs et calcule la complexité granule par granule.

Ainsi, on propose d'améliorer l'hypothèse d'indépendance en considérant qu'il est possible de « propager » le décalage fréquentiel de deux granules donnés au reste du motif, de sorte à évaluer une complexité différentielle et non plus absolue.

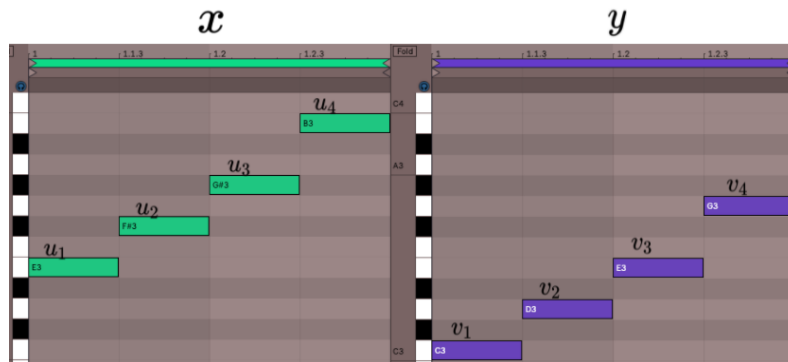


Figure 44: Exemple de deux motifs identiques à un décalage fréquentiel près.

Dans l'exemple ci-dessus (Fig. 44), x et y sont identiques, mais à un intervalle de 4 demi-tons près.

Si on effectue une translation globale de cet intervalle sur les notes des granules du motif antécédent, la complexité $\mathcal{C}(y|x)$ s'avère bien moindre que celle obtenue par un traitement granule par granule.

Ainsi, on propose deux méthodes de propagation des intervalles définis par les granules du couple de motifs.

Propagation du premier intervalle chromatique

Dans ce cas de figure, seul l'intervalle δ_1 correspondant aux notes du 1^{er} granule de chaque motif est propagé aux granules du motif antécédent.

$$\begin{aligned} \mathcal{C}(y|x) &\approx \mathcal{C}(v_1|u_1) + \mathcal{C}(v_2|u_1u_2v_1) + \mathcal{C}(v_3|u_1u_3v_1) + \mathcal{C}(v_4|u_1u_4v_1) \\ &= \mathcal{C}(\delta_1) + \mathcal{C}(\delta_2 - \delta_1) + \mathcal{C}(\delta_3 - \delta_1) + \mathcal{C}(\delta_4 - \delta_1) \end{aligned} \quad (3.15)$$

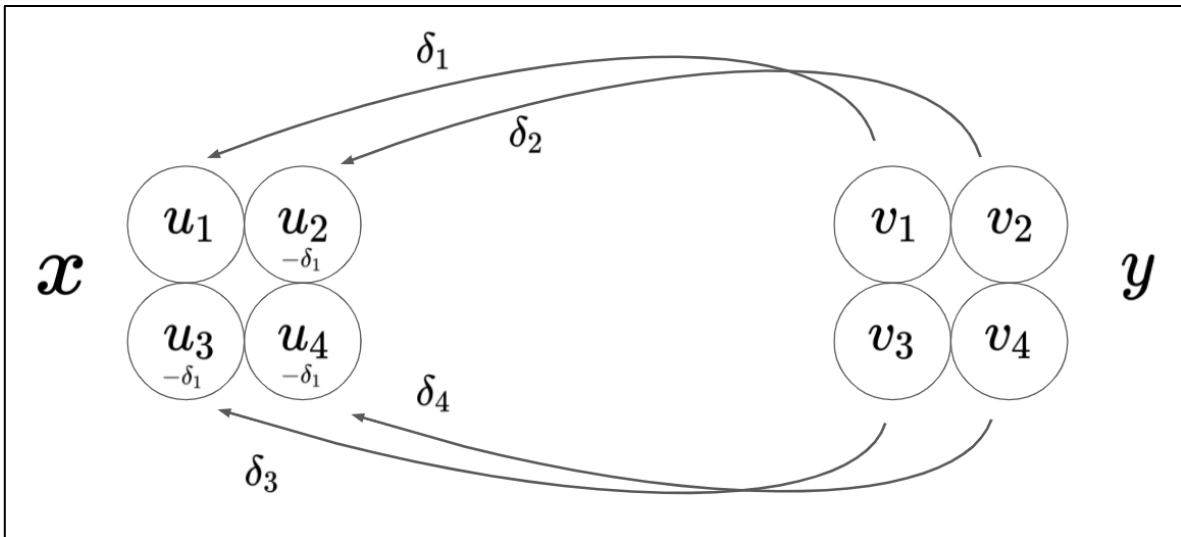


Figure 45: Représentation de l'estimation de la complexité $\mathcal{C}(y|x)$ selon l'hypothèse de propagation du premier intervalle δ_1 .

Propagation « en cascade » de chaque intervalle entre granules

Dans cette hypothèse, on se propose de ne pas seulement propager le premier intervalle chromatique déterminé entre les deux premiers granules, mais de propager chaque intervalle défini par chaque paire de granules, à tous les granules suivants dans l'ordre causal.

$$\begin{aligned} \mathcal{C}(y|x) &\approx \mathcal{C}(v_1|u_1) + \mathcal{C}(v_2|u_1u_2v_1) + \mathcal{C}(v_3|u_2u_3v_2) + \mathcal{C}(v_4|u_3u_4v_3) \\ &= \mathcal{C}(\delta_1) + \mathcal{C}(\delta_2 - \delta_1) + \mathcal{C}(\delta_3 - \delta_2) + \mathcal{C}(\delta_4 - \delta_3) \end{aligned} \quad (3.16)$$

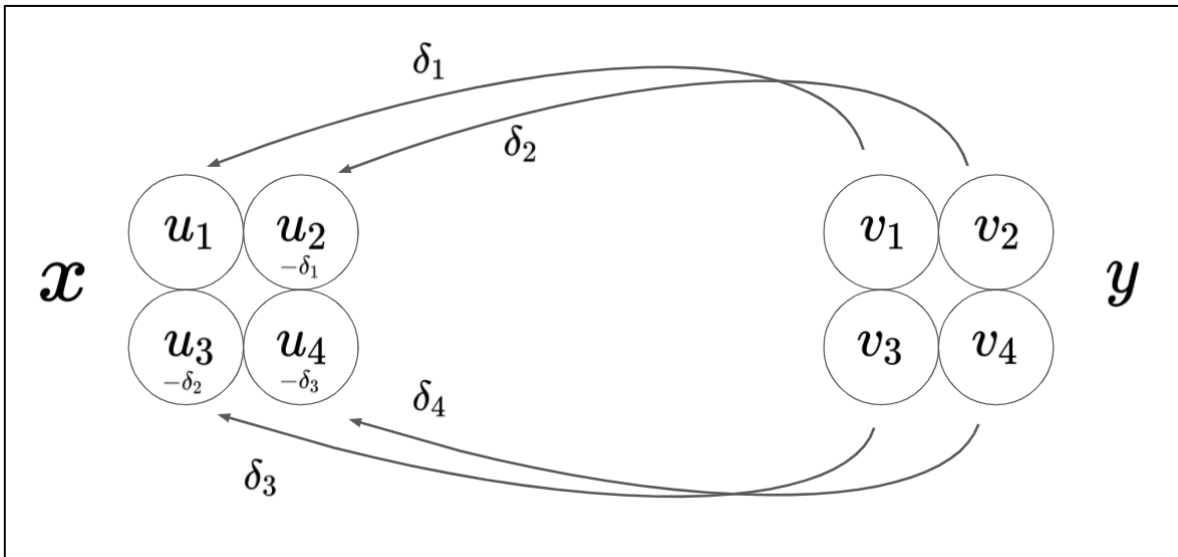


Figure 46: Expression de la complexité $\mathcal{C}(y|x)$ selon l'hypothèse de propagation d'intervalle de chaque couple de granules.

3.1.4.2 Recherche bilatérale

Lorsque l'on procède par propagation d'intervalles, les granules d'ancrage que l'on choisit conditionnent fortement la complexité (alors que cela n'a pas d'impact lorsque les granules sont considérés comme indépendants).

En effet, l'intervalle entre les notes des deux premiers granules est propagé au reste du motif et cet intervalle a un fort impact sur l'estimation de la complexité.

La méthode de base consiste à considérer que la référence du début du motif est définie par l'instant de début de son premier granule.

Néanmoins, on peut étendre le principe des méthodes précédentes en définissant le granule d'ancrage en tenant compte de la position du temps fort dans le motif.

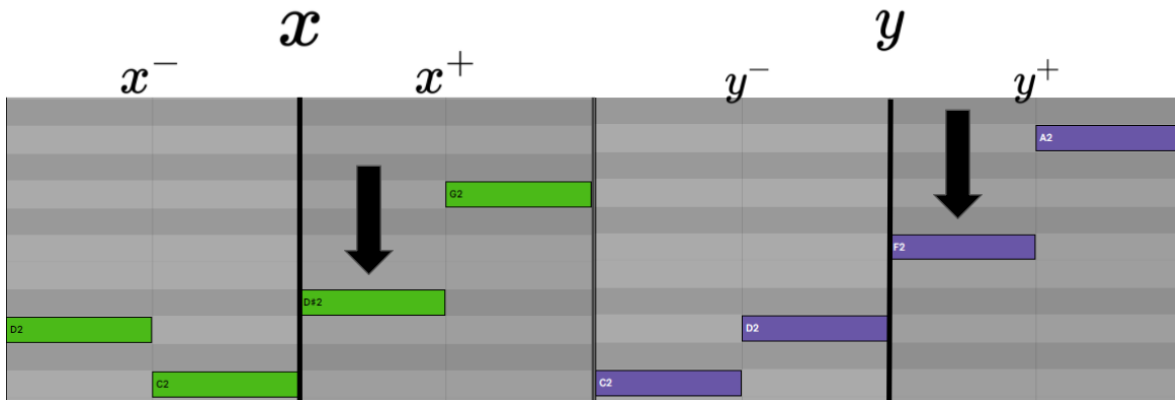


Figure 47: Deux motifs où le temps fort coïncide avec le début du granule 3.

Sur la figure précédente, on observe deux motifs pour lesquels le temps fort intervient au début du 3ème granule.

On repart de la définition d'un motif comme séquence de $m = 4$ granules notés jusqu'à présent :

$$x = u_1 u_2 u_3 u_4$$

Dans le cas de la recherche bilatérale, on définit dans ces 4 granules un granule de référence u_0 dont on considère qu'il sert d'ancrage au motif et on renumérote les autres granules par rapport à cette référence.

On réécrit alors x sous une des formes suivantes (selon la position de la référence) :

- $x = u_0 u_1 u_2 u_3$
avec $x^- = u_0$ et $x^+ = u_0 u_1 u_2 u_3$
 - $x = u_{-1} u_0 u_1 u_2$
avec $x^- = u_0 u_{-1}$ et $x^+ = u_0 u_1 u_2$
 - $x = u_{-2} u_{-1} u_0 u_1$
avec $x^- = u_0 u_{-1} u_{-2}$ et $x^+ = u_0 u_1$
 - $x = u_{-3} u_{-2} u_{-1} u_0$
avec $x^- = u_0 u_{-1} u_{-2} u_{-3}$ et $x^+ = u_0$
- (3.17)

Nous nous sommes contentés d'énumérer les cas pour les motifs de 4 granules, mais le principe peut se généraliser à des motifs d'un nombre quelconque de granules.

On note que, dans chaque cas, x est la concaténation de sa partie gauche x^- écrite en ordre rétrograde et de sa partie droite x^+ écrite en ordre direct, à ceci près que l'élément d'ancrage u_0 apparaît à la fois dans la partie gauche et droite.

Pour l'estimation de la complexité d'un motif y à partir de son antécédent x , on fait l'hypothèse que la position des granules d'ancrage dans x et y est identique du fait que le décalage est constant.

En toute généralité, sous l'hypothèse de recherche bilatérale, la complexité d'un motif y relativement à son antécédent x peut s'écrire comme :

$$\mathcal{C}(y|x) = \mathcal{C}(y^- |x^-) + \mathcal{C}(x^+ |y^+) - \beta \mathcal{C}(v_0|u_0) \quad (3.18)$$

Dans nos expériences, nous avons utilisé $\beta = 0$, ce qui revient à compter 2 fois la distance entre les granules d'ancrage.

Le calcul de l'estimation pour chacune des parties gauches $\mathcal{C}(y^- |x^-)$ et droites $\mathcal{C}(x^+ |y^+)$ s'effectue conformément aux principes décrits par l'hypothèse de propagation définie dans la section précédente.

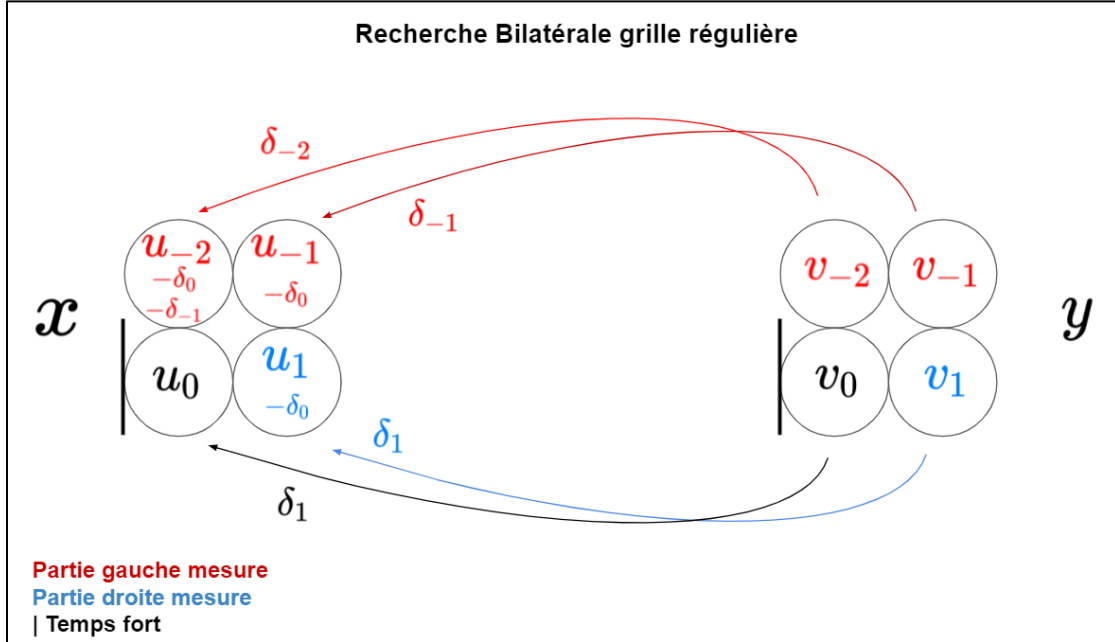


Figure 48: Illustration du principe d'estimation bilatérale avec propagation cascade

3.1.5 Distance entre deux granules

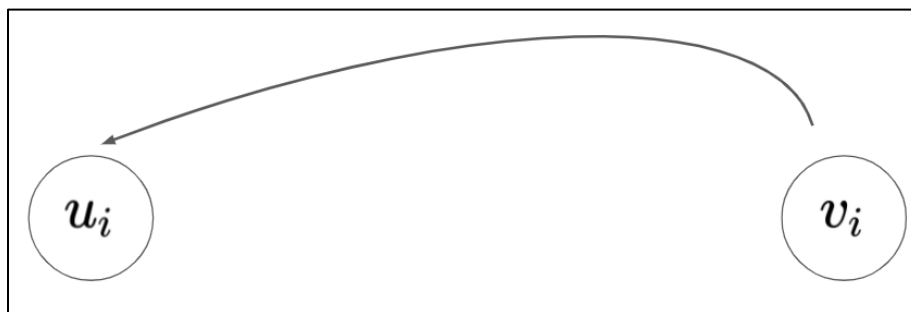


Figure 49: Distance entre deux granules

Nous arrivons maintenant à l'échelle la plus basse de la comparaison entre motifs : considérons deux granules u_i et v_i constitués chacun d'une note, lesquels correspondent donc à une durée de $\frac{1}{8}$ de mesure.

Alors que la distance entre notes sur une échelle chromatique semble la plus proche d'un transport minimal dans l'espace des intervalles, nous avons également considéré une autre métrique (dite « distance musicale ») qui tient compte de la corrélation des harmoniques entre notes et non pas de leur intervalle en termes de hauteur.

Plus précisément, en supposant une représentation d'une note en termes d'harmoniques uniformément réparties sur une large étendue, la distance musicale proposée correspond au transport d'harmoniques nécessaire pour amener une note sur l'autre.

Ainsi, il est beaucoup moins coûteux de transporter une note sur sa quinte (do > sol), sa quarte (do > fa) ou sa tierce majeure (do > mi) plutôt que sa seconde mineure (do > do#), sa septième majeure (do > si) ou sa quinte diminuée (do > fa#)

La figure 50 compare dans un même tableau les valeurs prises par la distance chromatique et la distance musicale sur les mêmes intervalles. Naturellement, les valeurs sont invariantes par changement du signe de la différence.

Mineurs, majeurs ou intervalles parfaits	Distance d_c	Distance d_m
Unisson	0	0
Seconde mineure	1	5
Seconde majeure	2	4
Tierce mineure	3	3.32
Tierce majeure	4	3
Quarte parfaite	5	2.58
Quinte diminuée	6	4.75
Quinte parfaite	7	2
Sixte mineure	8	3.58
Sixte majeure	9	2.81
Septième mineure	10	3.32
Septième majeure	11	4.46

Figure 50: Tableau récapitulatif des variantes de distances entre granules considérées dans cette thèse.
 d_c : distance chromatique – d_m : distance musicale

Dans les deux cas, on définit la distance entre deux granules u_i et v_i comme une fonction de l'intervalle chromatique δ entre leurs deux notes constituantes comme :

$$d(v_i|u_i) = d(\delta) \quad (3.19)$$

Cette distance peut se comprendre comme étant proportionnelle au nombre d'instructions élémentaires permettant de transformer u_i en v_i dans une représentation intervallique (distance chromatique) ou dans une représentation harmonique (redondance énergétique).

3.2 Inférence d'une segmentation optimale pour une méthode d'estimation donnée

3.2.1 Caractérisation d'une méthode algorithmique

On a présenté plusieurs approches pour estimer la complexité d'une séquence musicale s'appuyant sur différentes dépendances entre systèmes de 4 éléments, mais aussi des variantes de l'historique des motifs, différents schémas de propagation de transport entre granules, différentes définitions d'un granule d'ancrage ou encore différentes fonctions de transport entre deux granules.

Du fait de ces divers scénarios, on caractérise une méthode donnée par un ensemble Γ qui recense les 5 caractéristiques définissant ladite méthode :

$$\begin{aligned} \Gamma = \{ & \text{Critère de complexité entre systèmes,} \\ & \text{Critère de complexité interne à un système,} \\ & \text{Schéma de propagation du transport,} \\ & \text{Point d'ancrage des granules,} \\ & \text{Distance / coût de transport entre granules} \} \end{aligned} \quad (3.20)$$

Pour chacune de ces 5 caractéristiques, nous avons conservé, pour l'étude expérimentale, les 2 approches les plus performantes individuellement pour les tester en combinaison avec les autres caractéristiques, soit au total 32 méthodes qui forment un ensemble que nous noterons M .

La figure 51 résume sous forme d'un tableau les combinaisons de caractéristiques qui génèrent les 32 méthodes testées.

<i>Complexité entre systèmes</i>	
Indépendants	Prise en compte du Super-Système
<i>Complexité interne à un système</i>	
Motif « Primer » seul antécédent des autres motifs	Motif antécédent qui minimise la complexité du motif « Contraste ».
<i>Propagation du transport</i>	
Propagation de l'intervalle défini par les premiers granules de chaque motif	Propagation dite « cascade » des intervalles entre les granules
<i>Point d'ancrage des granules</i>	
Standard	Recherche bilatérale
<i>Distance élémentaire entre granules</i>	
Chromatique	Musicale

Figure 51: Table des caractéristiques des méthodes d'estimations, définissant l'ensemble M des méthodes Γ.

3.2.2 Détermination de θ^* et du score de complexité

Inférer la segmentation optimale selon une méthode d'estimation consiste à déterminer le décalage θ^* correspondant à la segmentation dont le score de complexité est le plus faible.

Le score d'une segmentation dépend donc du décalage θ par rapport à la mesure et du système de faces parallèles dans le tesseract, c'est-à-dire de la PPP π .

Pour une section X donnée, on a donc :

$$\theta^* = \text{ArgMin}_{\pi} \text{ArgMin}_{\theta} \mathbf{C}(X, \theta, \pi) \quad (3.21)$$

où \mathbf{C} représente la fonction de complexité $\mathbf{C}(X, \theta, \pi)$ de la segmentation de la section X segmentée en motifs avec un décalage θ et à laquelle on applique ensuite la PPP π . Si l'on procédait dans le sens inverse, l'estimation du décalage *après* permutation des mesures perdrait tout son sens.

On détaille dans cette section, les domaines de définition des variables θ et π et l'inférence des paramètres de segmentation optimaux θ^* et π^* .

Valeurs de décalage θ

Une segmentation de la section musicale est inférée par la définition du décalage θ de la grille de segmentation régulière par rapport à la première mesure de la section, le résultat correspondant étant noté X_{θ} .

Définir un décalage négatif par rapport à la première mesure permet de repérer les anacrouses ou anticipations du matériau musical. À l'inverse, un décalage positif permet de repérer les débordement ou effet de retard du matériau musical. Un décalage nul correspond à la segmentation de la section selon le découpage métrique.

Dans ces travaux, on considère que les décalages extrémaux sont de $\pm \frac{1}{2}$ mesure, correspondant à un décalage θ de ± 2 temps. Le pas élémentaire entre deux valeurs de décalage est défini comme $\frac{1}{8}$ de mesure, soit 0,5 temps.

On définit donc les différentes segmentations possibles selon le décalage de la façon suivante :

$$\begin{aligned} X_{\theta} &= x_1^{\theta} \dots x_t^{\theta} \dots x_z^{\theta} \\ X_{\theta} &= [u_{\theta+1} \dots u_{\theta+4}] \dots [u_{4d+\theta+1} \dots u_{4(d+1)+\theta}] \dots [u_{(g-3)+\theta} \dots u_{g+\theta}] \end{aligned} \quad (3.22)$$

$\theta \in [-2, 2]$ temps avec un pas de 0,5 temps

On aboutit ainsi à 9 valeurs de décalage à considérer pour chaque extrait traité.

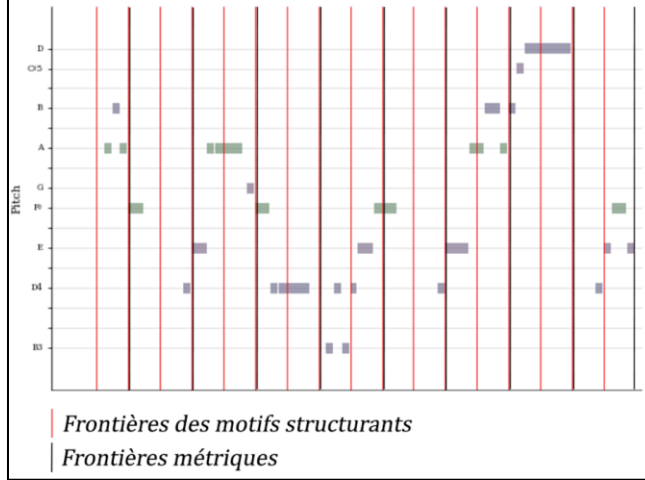


Figure 52: Exemple de segmentation selon la grille régulière et un décalage θ de -2 temps.

Systèmes de faces parallèles π

Chaque décalage θ donne lieu à un ensemble de 16 motifs pour la section musicale segmentée X_θ .

Comme discuté dans la section 2.1.4, on étudie les relations qu'entretiennent les couples de motifs à plusieurs échelles temporelles différentes.

Sur le tesseract, on peut regrouper ces 16 motifs en 4 systèmes d'implications (constitués de 4 motifs chacun).

Comme on l'a explicité dans le chapitre précédent, il existe 6 façons différentes de définir ces 4 systèmes, correspondant aux 6 systèmes de faces parallèles dans le tesseract, eux-mêmes associés à une fonction de permutation particulière PPP nommée π_p (avec donc $p \in [0, 5]$). On peut donc écrire la permutation π_p de la segmentation X^θ sous la forme :

$$X^\theta[p] = S_1^\theta[p] S_2^\theta[p] S_3^\theta[p] S_4^\theta[p] \quad (3.23)$$

Avec ces nouvelles notations, et compte tenu des domaines de définition choisis, on peut donc réécrire l'équation 3.2.

$$\theta^* = \text{ArgMin}_{p \in [0,5]} \text{ArgMin}_{\theta \in [-2,2]} \mathbf{C}(X^\theta[p]) \quad (3.24)$$

Et on a donc, pour une fonction de complexité donnée, $6 \times 9 = 54$ possibilités à tester pour chaque extrait X .

Détermination des paramètres de segmentation optimaux θ^* et π^*

Pour une méthode Γ donnée, compte tenu du relativement faible nombre de possibilités pour $X^\theta[p]$, la fonction de calcul de complexité d'une section musicale est exécutée pour chaque couple (θ, p) ce qui revient à une approche de type « *force brute* ».

La détermination de la segmentation optimale θ^* s'effectue en recherchant parmi ces 54 possibilités, le score de complexité minimal.

En cas d'égalité, on conserve la valeur de θ la plus régulière au sens métrique, le décalage nul sera donc prioritaire, suivi des décalages de -2 et 2 temps, puis les décalages -1 et 1, etc..

On retourne alors cette valeur de complexité minimale ainsi que l'ensemble associé $\{\theta^*, p^*\}$ correspondant aux paramètres de segmentation optimaux de la section X .

Pour une méthode Γ donnée, on notera ces estimations $\theta^*(\Gamma)$ et $p^*(\Gamma)$.

3.2.3 Calcul du score de complexité d'une segmentation

Dans sa forme conditionnelle, la complexité C que nous utilisons dans ce travail s'interprète comme la quantité d'information (en nombre de bits) nécessaire pour compresser une nouvelle information par rapport à un historique, dans un paradigme de complexité algorithmique, selon lequel un événement est d'autant plus économique à compresser qu'il est simple à décrire (en terme d'instructions algorithmiques élémentaires) connaissant son contexte.

Une fois que l'on a déterminé le décalage θ et le système de faces parallèles π_p , on peut estimer la complexité de la section musicale $X_\theta[p]$.

Cette complexité dépend de l'ensemble des complexités des objets musicaux qui la constitue, ainsi que des approches utilisées dans la méthode Γ .

Comme présenté dans la section 3.1, les objets musicaux utilisés pour le calcul de la complexité sont imbriqués les uns dans les autres, ce qui se traduit par l'enchâssement des traitements correspondants dans l'architecture de l'algorithme.

En effet, la complexité d'une section dépend de la complexité de chacun de ses systèmes, qui dépendent de la complexité de leurs motifs constituants, eux-mêmes dépendants du coût algorithmique du transport entre leurs granules.

Le schéma de la figure 53 ci-dessus illustre l'architecture de la fonction de calcul de complexité, où chaque appel de fonction conduit au calcul des quantités correspondant aux équations détaillées dans la section 3.1.

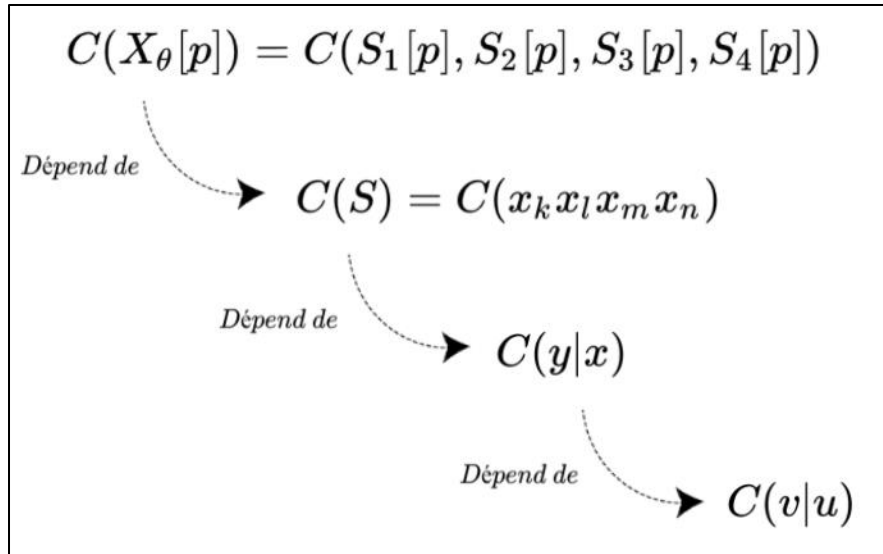


Figure 53: Schéma explicatif du calcul de complexité avec dépendances entre objets musicaux à 4 échelles temporelles.

3.2.4 Combinatoire, temps de calcul et optimisation

Dans cette section, on apporte quelques détails liés à la combinatoire, au temps d'exécution et aux optimisations possibles de l'algorithme de segmentation.

La table suivante présente le nombre d'appels de chaque fonction d'estimation de complexité d'une section musicale pour une méthode algorithmique Γ donnée.

Certaines méthodes d'estimation provoquent des appels supplémentaires de fonction à certaines échelles temporelles, notamment dans le cas des approches par « prise en compte du super-système », par « recherche du motif antécédent qui minimise la complexité du contraste » et par « recherche bilatérale ».

Echelles musicales	Nombres d'appels (en fonction de Γ)
Section	1
Système	4 ou 5
Motif	3 ou 5
Granules	4 ou 5

Pour une méthode Γ , un décalage θ et un système de face parallèle π donnés, la fonction d'estimation de complexité d'un granule par rapport à son antécédent est invoquée :

- au minimum : $1 \cdot 4 \cdot 3 \cdot 4 = 48$ fois
- au maximum : $1 \cdot 5 \cdot 5 \cdot 5 = 125$ fois

Pour chaque méthode Γ , 9 valeurs de décalages θ sont possibles, ainsi que 6 systèmes de faces parallèles π .

Dans le cas d'une méthode Γ qui maximise le nombre d'appels, la fonction $\mathcal{C}(v|u)$ de l'estimation de complexité à l'échelle granulaire est appelée 6750 fois pour une estimation de segmentation optimale.

Le temps d'exécution d'une estimation de segmentation selon la méthode Γ qui maximise le nombre d'appels est de l'ordre de 10 secondes sur une machine « Core i7 7th Gen » avec 16Go de RAM et utilisée avec Windows 10.

Du fait de ce faible temps de calcul, nous n'avons pas cherché à optimiser cet algorithme.

Néanmoins, il est possible de minimiser la quantité de calcul effectuée, du fait que certaines estimations de complexités conditionnelles de motifs sont communes à plusieurs des fonctions PPP.

On pourrait donc pré-tabuler l'ensemble des calculs d'estimations entre motifs et estimer la complexité d'une PPP en faisant la somme d'un sous-ensemble de valeurs issues de cette table, retournées à partir des indices des motifs estimés.

Inutile dans le cas présent, ce type d'optimisation peut s'avérer valable dans le cas où l'on travaillerait sur d'importantes bases de données ou des polytopes d'échelles largement supérieures.

3.2.5 Détails d'implémentations

Lecture de fichier MIDI

Le format MIDI est un format de données musicales très commun.

Ce format correspond à l'équivalent numérique d'une partition dans un format néanmoins différent. On y retrouve les informations concernant les instants auxquels les notes sont jouées, leur hauteur et vélocité, la signature temporelle, éventuellement le tempo ainsi que des informations concernant l'instrumentation.

Les développements algorithmiques de ces travaux de thèse ont été effectués en langage Python et en grande partie à l'aide du module Music21 créé par Michael Cuthbert (Cuthbert, 2010). Ce module est largement utilisé en MIR et constitue une ressource très utile pour l'étude musicologique algorithmique.

Music21 permet notamment la lecture et la conversion de fichier MIDI en objet de type Stream, qu'il est possible d'analyser et d'éditer facilement en Python.

Cette instanciation de la section musicale en Stream permet l'accès aux informations présentes dans le fichier MIDI et notamment les attributs des notes : leur hauteur, leur instant d'activation, leur durée, etc.

Ces objets et attributs sont manipulés dans les travaux de cette thèse pour la détermination de segmentation optimale et la génération des recompositions polytopiques.

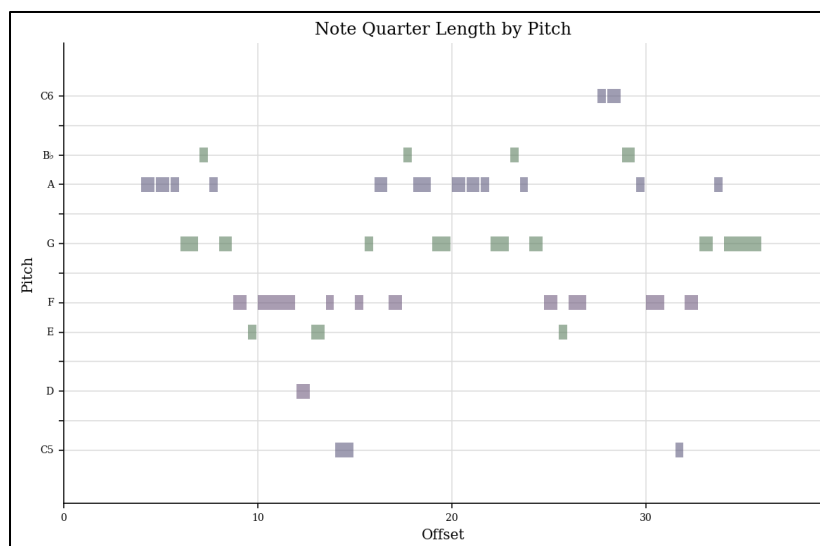


Figure 54: Exemple d'affichage des notes d'un fichier MIDI avec music21.

Réduction de la section musicale

Comme détaillé dans la section 2.1.1, l'information MIDI brute passe, préalablement à son traitement par nos algorithmes, par un processus de réduction qui consiste en une simplification et un échantillonnage de la mélodie en ensemble de g éléments granulaires.

On rappelle que notre analyse porte sur les refrains des lignes mélodiques et dont la surface musicale est de durée 8 mesures en signature temporelle 4/4.

Toutefois, afin de pouvoir déterminer les décalages possibles θ de la surface musicale par rapport à la première barre de mesure, on conserve (en plus des 8 mesures définies par les frontières métriques) le matériau musical sur $\frac{1}{2}$ mesure avant la première barre de mesure, ainsi que $\frac{1}{2}$ mesure après la fin de la 8^{ème} mesure. Les données manipulées pour la détermination du décalage θ sont donc des successions de 9 mesures en signature temporelle 4/4 et ont donc une durée de 36 temps.

L'opération de réduction consiste à échantillonner cette section musicale à une fréquence de $\frac{1}{8}$ de mesure, en un ensemble de 72 granules, à partir desquels on travaille ensuite sur une fenêtre glissante de 64 granules. Les notes dont l'attaque se produit entre les instants d'échantillonnages ne sont pas prises en compte dans la version réduite, mais remplacées par la note active à l'instant d'échantillonnage antérieur.

On opère ainsi une réduction de l'information brute dans le sens où les instants de début des notes et leurs durées sont recalées sur une grille métrique avec $\frac{1}{8}$ de mesure pour pas. Dans les logiciels dits de station de travail audio numérique (DAW), ce processus de recalage est appelé la *quantification*, mais il s'agit en fait plutôt d'un échantillonnage...

3.3 Méthodes algorithmiques et évaluation préliminaire

Dans cette section on présente une méthodologie permettant d'évaluer les méthodes d'estimation par étude de leur concordance avec des *annotations* produites manuellement par 2 annotateurs.

3.3.1 Génération des recompositions polytopiques

On présente dans cette section l'algorithme permettant de générer les recompositions polytopiques définies par les fonctions PPP.

Dans les sections précédentes, les fonctions PPP définies par les permutations π_p ont été utilisées afin de déterminer le système de faces parallèles qui apporte la valeur optimale de complexité de la section musicale.

Ces fonctions sont également utilisées lors de la phase de recomposition, en tant que permutations appliquées à l'ensemble des indices des motifs afin d'en modifier l'ordre temporel et produire les variations musicales polytopiques.

Cette étape de génération des recompositions intervient une fois que l'on a fixé (manuellement ou par méthode algorithmique) un décalage θ^* pour segmenter la section X .

À la différence de l'algorithme d'estimation de la complexité, nous n'avons cependant pas besoin ici de procéder à la *réduction* de la section musicale, du fait qu'il n'y a pas d'analyse granulaire en jeu. On veut, au contraire, conserver au maximum les informations musicales de l'extrait original, y compris les offsets des notes.

Il est toutefois nécessaire de procéder à un léger « recalage » des notes à l'échelle granulaire de la grille de segmentation ($\frac{1}{8}$ mesure) afin que les motifs structurants soient délimités sans introduire d'artefacts.

Après lecture et conversion du fichier MIDI de la section musicale d'origine, on instancie les 16 motifs structurants selon la grille de segmentation et le décalage θ^* .

L'opération de segmentation peut avoir pour effet de scinder des notes si elles sont prolongées au-delà de la frontière d'un motif à l'autre. Un processus de « *recollage* » est donc prévu après permutation, dans le cas où deux notes identiques scindées se retrouvent à nouveau juxtaposées dans la recomposition.

Une fois les motifs structurants x_1, x_2, \dots, x_{16} instanciés par la segmentation, on applique chacune des 6 fonctions de permutations π_p ($0 \leq p \leq 5$) à cet ensemble afin de modifier l'ordre temporel des motifs structurants tout en respectant les systèmes de faces dans tesseract.

3.3.2 Annotations : segmentations manuelles par deux personnes du domaine

Afin d'évaluer, dans une phase préliminaire, les résultats obtenus par les méthodes d'estimation algorithmique, deux personnes travaillant dans le domaine de l'informatique musicale (en l'occurrence, l'auteur de cette thèse et son directeur de thèse !) ont eu pour tâche de segmenter manuellement un ensemble de 52 refrains du corpus RWC POP, sur lesquels portent les expériences. Ces deux annotateurs ont travaillé séparément sans se concerter, comme en atteste le faible score de concordance de leur résultat (moins de 35 % d'accord inter-annotateur !)

Comme les segmentations algorithmiques, les segmentations manuelles sont effectuées selon une grille de segmentation régulière, autrement dit, il s'agit pour l'annotateur de déterminer un décalage θ^* « optimal ».

Il convient de préciser qu'afin de faciliter cette tâche d'annotation, on a développé une interface interactive en langage Processing.

Cette interface permet la visualisation et la navigation parmi les fichiers MIDI des sections musicales d'origine, la visualisation des frontières de la grille de segmentation et le décalage de cette grille par variation du paramètre θ .

Cette interface permet également, après avoir instancié la segmentation voulue, d'écouter la version originale ainsi que chacune des 5 recompositions polytopiques générées depuis la segmentation définie.

Cela permet donc aux annotateurs de déterminer facilement le décalage θ^* qu'ils considèrent comme optimal sur la base des recompositions obtenues. Les décalages sont donc optimisés à l'écoute du résultat et non pas seulement sur des critères indirects prenant en compte les propriétés musicologiques de l'original.

Deux jeux d'annotations manuelles indiquant le décalage optimal sur 52 refrains de la base RWC POP ont ainsi été collectés, et ces annotations de référence sont notées A_1 et A_2 .

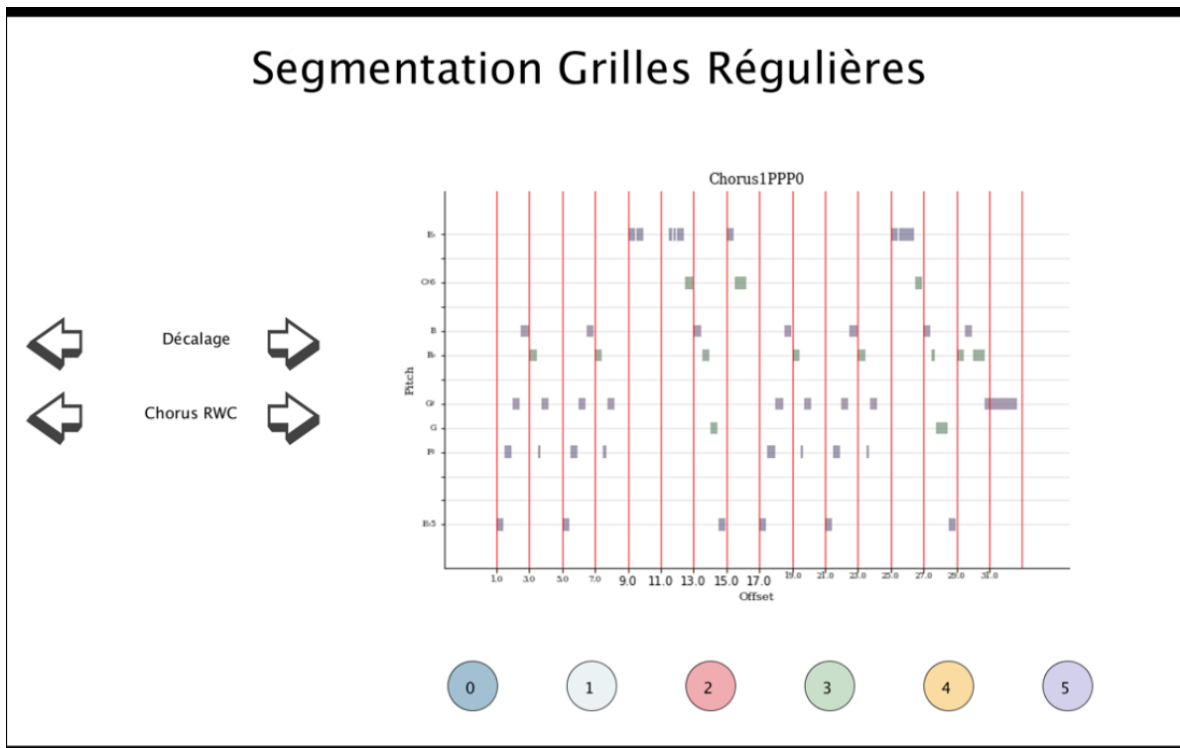


Figure 55: Interface de segmentation manuelle pour la collecte des annotations de référence. Les boutons colorés permettent d’écouter les recompositions polytopiques selon la grille choisie.

3.3.3 Evaluation par rapport aux annotations manuelles

Comme présenté dans la partie 3.1.1, une méthode d’estimation est définie par 5 caractéristiques binaires.

Dans cette section, on présente une méthodologie ayant pour but d’évaluer une méthode d’estimation en fonction de la corrélation de son résultat par rapport à un ou plusieurs jeux d’annotations manuelles.

Deux estimations de concordance sont développées dans ce chapitre, on commence par détailler la façon d’évaluer la concordance d’une méthode algorithmique avec un jeu d’annotation manuelle.

Par la suite, on discutera une méthode permettant d’évaluer la concordance d’un sous-ensemble d’approches, correspondant à une des caractéristiques des méthodes d’estimation Γ .

Score de concordance avec les annotations de référence

Pour notre ensemble d’étude constitué de $R = 52$ refrains, les jeux d’annotations manuelles de référence A_1 et A_2 , sont deux ensembles constitués de 52 décalages θ^* « optimaux » déterminés manuellement pour les 52 refrains de cette étude.

Ils sont notés comme suit :

$$\Theta(A_1) = \{\theta_i^*(A_1)\}_{1 \leq i \leq R} \quad (3.25)$$

$$\Theta(A_2) = \{\theta_i^*(A_2)\}_{1 \leq i \leq R} \quad (3.26)$$

Soit maintenant Γ_k une configuration de caractéristiques pour une approche algorithmique donnée ($k \in [1, 32]$). On définit $\theta_i^*(\Gamma_k)$ l'ensemble des décalages correspondant aux valeurs optimales établies algorithmiquement par la méthode Γ_k sur chacun des 52 refrains du corpus de l'étude.

$$\Theta(\Gamma_k) = \{\theta_i^*(\Gamma_k)\}_{1 \leq i \leq R} \quad (3.27)$$

On définit alors le score de concordance entre les annotations $\Theta(A_j)$ et $\Theta(\Gamma_k)$ comme suit :

$$T_{jk} = \frac{\sum_{i=1}^R \tau_{ijk}}{R} \quad (3.28)$$

avec $\tau_{ijk} = 1$ si $\theta_i^*(A_j) = \theta_i^*(\Gamma_k)$ et $\tau_{ijk} = 0$ sinon

Dans les résultats rapportés, la valeur de T_{jk} est exprimée en pourcentage et correspond à la proportion de résultats identiques entre la configuration d'estimation et le jeu d'annotations manuelles.

Dans le chapitre 2, on a vu que l'estimation de la complexité était très contrainte par les hypothèses définies par les caractéristiques des méthodes d'estimations Γ considérées.

Un total de $2^c = 32$ (donc $c = 5$) méthodes d'estimations différentes sont étudiées, pour chaque section musicale X on obtient donc 32 résultats de segmentation $\{\theta^*(\Gamma), \pi^*(\Gamma)\}$.

Il peut arriver que plusieurs méthodes aient le même décalage optimal θ^* , ce qui a comme effet de produire les mêmes recompositions polytopiques. Cette propriété sera utilisée dans la phase d'évaluation perceptive présentée dans le chapitre suivant.

Score de discordance entre variantes de caractéristiques

Par ailleurs, on souhaite calculer des scores qui permettent de départager 2 variantes d'une même caractéristique des méthodes d'estimation, indépendamment des autres caractéristiques : par exemple, pour déterminer s'il est préférable de rajouter le terme de complexité du Super-Système ou pas, indépendamment des autres caractéristiques des méthodes.

On va donc étudier les différentes méthodes, *caractéristique par caractéristique*, en les appariant, c'est-à-dire en les couplant 2 par 2 lorsqu'elles ne diffèrent que par la seule caractéristique d'intérêt. On rappelle que l'on a 2^c méthodes où $c = 5$ est le nombre total de caractéristiques des méthodes.

	Méthode N°1	Méthode N°9
Complexité entre systèmes	<i>Super – Système</i>	<i>Systèmes Indépendants</i>
Complexité dans un système	<i>Recherche de l'antécédent du motif contraste</i>	<i>Recherche de l'antécédent du motif contraste</i>
Recherche	<i>Bilatérale</i>	<i>Bilatérale</i>
Propagation	<i>Cascade</i>	<i>Cascade</i>
Distance	<i>Musicale</i>	<i>Musicale</i>
Concordance à A_1	$T_{11} = 23,1\%$	$T_{19} = 26,9\%$
Concordance à A_2	$T_{21} = 26,9\%$	$T_{29} = 30,8\%$

Figure 56: Exemple de 2 méthodes appariées avec une différence sur leur 1^{ère} caractéristique.

Dans le cas de figure de l'exemple donné en figure 57, les 2 configurations font partie de l'ensemble des méthodes appariées pour l'étude de la "*Complexité entre systèmes*" du fait que l'hypothèse dissemblable concerne cette caractéristique uniquement.

Pour une méthode d'estimation Γ donnée, on définit donc par un vecteur de booléens $(b_\ell^\Gamma)_{1 \leq \ell \leq c}$ le statut de la variante d'une caractéristique ℓ : 0 ou 1 selon qu'elle est absente ou présente dans la méthode Γ .

Pour la caractéristique ℓ on a donc $Q = 2^{c-1}$ couples de méthodes appariées

Afin de départager deux variantes pour une caractéristique ℓ donnée par rapport à une annotation de référence A_j , on calcule la différence de concordance entre les couples de méthodes avec variante d'une part et sans variante d'autre part.

Deux méthodes appariées quelconques seront notées Γ_q^ℓ et $\Gamma_q^{\bar{\ell}}$, avec $1 \leq q \leq Q$.

Deux scores de discordance Y_1 et Y_2 sont utilisés pour mesurer le degré de divergence des résultats donnés par les méthodes appariées vis-à-vis du jeu d'annotation A_j .

- Y_{1j} mesure le nombre de fois où l'une des variantes à l'avantage sur l'autre sur l'ensemble des méthodes selon la référence A_j .

$$Y_{1j} = \sum_1^{c-1} \text{sgn} (T_{jk}^\ell - T_{jk}^{\bar{\ell}}) \quad (3.29)$$

où sgn désigne la fonction signe avec la convention $\text{sgn}(0) = 0$ et où T_{jk}^ℓ (resp. $T_{jk}^{\bar{\ell}}$) désigne le score de concordance de la section précédente restreint aux variantes présentant (resp. ne présentant pas) la caractéristique ℓ , c'est-à-dire $\{\Gamma_q^\ell\}$ et $\{\Gamma_q^{\bar{\ell}}\}$.

- Y_{2j} mesure la somme des différences de pourcentages de concordance des méthodes appariées avec le jeu d'annotations A_j . Avec les mêmes notations que pour l'équation (3.10) :

$$Y_{2j} = \sum_1^{c-1} (T_{jk}^\ell - T_{jk}^{\bar{\ell}}) \quad (3.30)$$

qui tient compte de la valeur effective de l'écart entre les concordances et pas simplement leur signe.

On calcule par la suite:

$$\bar{Y}_1^\ell = \frac{1}{a} \sum_{j=1}^a Y_{1j} \quad (3.31)$$

$$\text{et } \bar{Y}_2^\ell = \frac{1}{a} \sum_{j=1}^a Y_{2j} \quad (3.32)$$

rendant compte de la préférence moyenne de l'ensemble des a annotateurs pour la caractéristique ℓ selon les deux mesures Y_1 et Y_2 .

On détermine finalement la caractéristique ℓ^* ayant la meilleure concordance par rapport aux a annotateurs par :

$$\ell^* = \text{Argmax}_\ell (| \bar{Y}_1^\ell |) \quad (3.33)$$

En cas d'égalité, la caractéristique ℓ^* est arbitrée par :

$$\ell^* = \text{Argmax}_\ell (| \bar{Y}_2^\ell |) \quad (3.34)$$

Ainsi, on détermine la caractéristique ℓ^* permettant d'obtenir le meilleur « consensus » parmi l'ensemble des a annotateurs.

L'ensemble des scores de concordance / discordance définis dans cette section et la section précédente ne procure pas une évaluation déterminante, du fait de la subjectivité des annotations.

Toutefois, ces scores vont permettre de guider un processus de sélection des méthodes à tester dans le cadre de l'évaluation perceptive qui fait l'objet du chapitre 4.2.

3.3.4 Résultats préliminaires

Une analyse préliminaire visant à estimer la concordance des différentes méthodes d'estimation de décalage avec les deux jeux d'annotations manuelles A_1 et A_2 a été menée.

Les deux indicateurs de discordance définis dans par les équations 3.29 et 3.30 permettent de comparer les méthodes d'estimations de décalage, caractéristique par caractéristique.

32 méthodes d'estimations de décalage étant considérées, on a un ensemble de 16 méthodes appariées pour chacune des 5 caractéristiques.

On calcule les deux indicateurs de discordance Y_1 et Y_2 définis dans la section 3.3.3 pour chaque caractéristique et pour chacun des deux jeux d'annotations de référence.

On obtient la table suivante :

	Indep/SuperSys	Primer/MinContr.	Standard/Bilat.	1 ^{er} Granule /Cascade	Musicale/Chroma
Y_{1j}					
$A_1 (j = 1)$	1	5	-7	3	12
$A_2 (j = 2)$	-5	10	0	-11	7
Y_{2j}					
$A_1 (j = 1)$	0,18	0,56	-0,18	0,06	1,32
$A_2 (j = 2)$	-0,33	0,83	0,07	-0,87	0,75
\bar{Y}_1^ℓ					
	-2	7,5	-3,5	-4	9,5
\bar{Y}_2^ℓ					
	-0,075	1,035	-0,055	-0,405	0,695

Figure 57: Table des résultats de l'évaluation de concordance des caractéristiques algorithmiques

On constate que la caractéristique de « Distance » donne la valeur maximum de \bar{Y}_1^ℓ (en valeur absolue), c'est donc cette caractéristique ℓ^* qui permet de faire le meilleur consensus entre les annotateurs.

D'autre part, cette valeur est positive, ce qui indique que la caractéristique de « Distance musicale » l'emporte sur la caractéristique de « Distance chromatique ».

Sur cette base, on peut donc fixer la caractéristique « Distance » au statut de « Distance musicale » qui procure une concordance beaucoup plus forte avec les annotations que le statut de « Distance chromatique ».

En conséquence, toutes les méthodes avec la caractéristique « Distance chromatique » sont écartées ce qui a pour effet de réduire de moitié l'ensemble des méthodes qu'il reste à considérer.

La table suivante montre les configurations restantes :

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4 Systèmes	Sup.Sys.	Sup.Sys.	Sup.Sys.	Sup.Sys.	Sup.Sys.	Sup.Sys.	Sup.Sys.	Sup.Sys.	Sys. Indép.	Sys. Indép.	Sys. Indép.	Sys. Indép.	Sys. Indép.	Sys. Indép.	Sys. Indép.	Sys. Indép.
Estimation 1 Système	Min.Contrast	Min.Contrast	Min.Contrast	Min.Contrast	Primer	Primer	Primer	Primer	Min.Contrast	Min.Contrast	Min.Contrast	Min.Contrast	Primer	Primer	Primer	Primer
Estimations Motif	Bilat.	Bilat.	1st	1st	Bilat.	Bilat.	1st	1st	Bilat.	Bilat.	1st	1st	Bilat.	Bilat.	1st	1st
Propagation	Casc. All	Prop. 1st	Casc. All	Prop. 1st	Casc. All	Prop. 1st	Casc. All	Prop. 1st	Casc. All	Prop. 1st	Casc. All	Prop. 1st	Casc. All	Prop. 1st	Casc. All	Prop. 1st
Distance	Mus															

Figure 58: Configuration restantes après élimination de la caractéristique « distance chromatique » entre granules.

Analysons maintenant le score de concordance de ce nouvel ensemble de 16 méthodes avec chacun des jeux d'annotations selon le paramètre T défini dans la section 3.3.3.

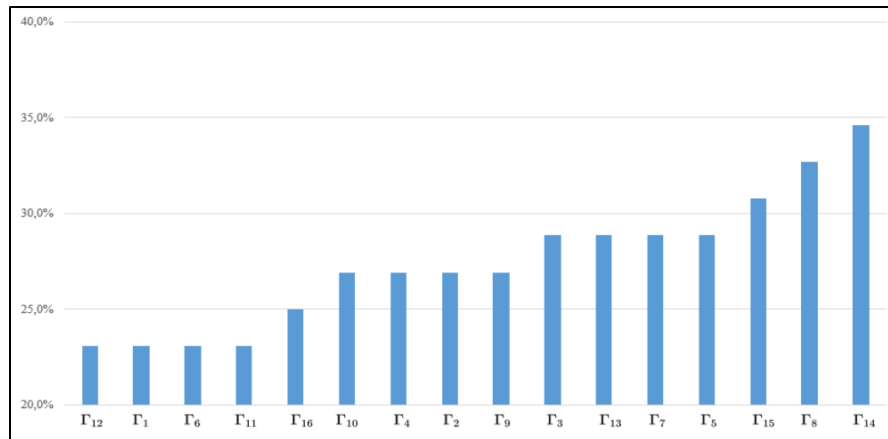


Figure 59 : Concordance des méthodes algorithmiques avec le jeu d'annotation A₁

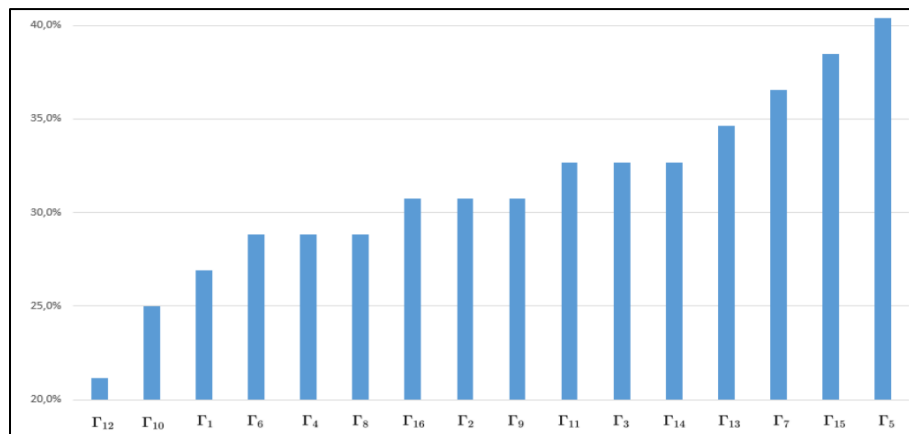


Figure 60 : Concordance des méthodes algorithmiques avec le jeu d'annotation A₂

On constate d'après ces 2 graphiques qu'il n'y pas de tendance véritablement commune des scores de concordance des méthodes d'estimation de décalage entre les deux jeux d'annotations.

Ces résultats ne permettent donc pas de déterminer de façon déterminante une méthode d'estimation d'offset algorithmique qui rende compte des préférences des deux annotateurs.

Dans la suite de ce travail, on utilisera une méthodologie « d'élimination successive » (section 4.2.3) pour guider un compromis entre les préférences des deux jeux d'annotations.

Du fait qu'il soit difficile de déterminer une tendance claire pour départager la pertinence de cet ensemble de méthode, on décrit dans le chapitre suivant comment un sous-ensemble particulier de 6 méthodes ont été sélectionnées pour être ensuite évaluées perceptivement par un panel d'auditeurs.

Chapitre 4

Tests perceptifs

Les expériences de la section précédente mettent en évidence les difficultés qui existent pour évaluer objectivement des recompositions polytopiques et les comparer :

- 1) Il n'existe pas à ce jour de « vérité-terrain » objective pour déterminer de façon non-ambigüe ce qu'est une « bonne » décomposition polytopique.
- 2) Différents annotateurs aboutissent à des annotations différentes. Dans le cas présent, la concordance entre les deux annotations n'atteint que 34,6 % !
- 3) De multiples variantes de méthodes de segmentation sont envisageables et sont difficiles à départager car aucune ne concorde de façon décisive avec les annotations.

C'est pourquoi nous avons entrepris d'explorer la pertinence des recompositions polytopiques au moyen de protocoles d'évaluation perceptive consistant à compléter les mesures de performance algorithmiques par des jugements subjectifs de panels d'auditeurs.

Dans ce chapitre, on présente la méthodologie et les résultats de deux études expérimentales basées sur des populations d'auditeurs à l'aide de tests perceptifs.

Plusieurs buts sont visés à travers ces études.

- **Test 1** : dans une première série de tests, on mesure perceptivement l'acceptabilité musicale de recompositions polytopiques dont la segmentation a été effectuée manuellement, par rapport à des recompositions aléatoires ne vérifiant pas les critères d'invariance des recompositions polytopiques. On s'intéresse aux dégradations relatives induites par les différentes fonctions de recomposition PPP vs des recompositions quelconques.
- **Test 2** : dans une seconde étude, on évalue perceptivement l'acceptabilité musicale de recompositions polytopiques générées à partir de sections musicales segmentées par différentes méthodes algorithmiques d'estimation du décalage θ . Et compte tenu du grand nombre de méthodes en compétition, on procède préalablement à la sélection d'un sous-ensemble de méthodes candidates supposément représentatives de l'intégralité des méthodes à tester (le protocole de sélection utilisé est décrit dans la suite de ce chapitre).

4.1 Test 1 : évaluation des recompositions polytopiques vs aléatoires

4.1.1 Données

Pour cette expérience, nous avons utilisé le corpus RWC-POP, un sous-ensemble de la base de données Real World Computing (RWC) créée à des fins scientifiques par Goto et al. (Goto, 2002). Cette base de données a été conçue pour le domaine de la recherche et est disponible à faible coût sans restriction.

L'ensemble RWC POP est constitué de 100 morceaux au format WAV et MIDI. D'après les auteurs, ces morceaux ont été composés en partie dans le style des hits américains des années 80 et en partie dans le style de la pop japonaise des années 90.

Ce corpus a été largement utilisé pour diverses tâches dans le domaine de la recherche d'information musicale (Music Information Retrieval).

Un autre avantage à utiliser les corpus RWC-POP est qu'il n'y a potentiellement aucun risque à ce que les sujets soient déjà familiers avec les morceaux présentés. Cette situation est souhaitable, car le fait de connaître un extrait musical préalablement pourrait affecter la perception de la validité de leur recomposition d'une façon difficile à évaluer.

Pour cette expérience, un sous-ensemble de 24 morceaux a été sélectionné et plus particulièrement le premier refrain sur exactement huit mesures. La durée moyenne des extraits sélectionnés est de 19,5 secondes.

Dans le cadre de ce test, nous conservons la ligne mélodique, l'accompagnement harmonique, la basse et la ligne de batterie.

Chacune des parties instrumentales est segmentée manuellement en 16 motifs structurants d'après les principes présentés dans les chapitres précédents et les motifs sont permutés de façon synchrone lors de l'application de fonctions de permutations. Mais seule la mélodie subit éventuellement un décalage de surface (l'harmonie et le rythme restant synchronisés sur la métrique d'origine).

10 fonctions de permutations différentes sont appliquées à la section musicale d'origine, les 5 fonctions PPP (présentées dans le chapitre 2.3.2), ainsi que 5 fonctions de permutations « aléatoires » dites *Randomly Generated Permutations* (RGP).

Les permutations RGP sont sélectionnées en conservant les mêmes propriétés topologiques (voir section 2.3.3) que les fonctions PPP et induisent ainsi les mêmes degrés de discontinuité et d'excursion. Tout comme les PPP, le 1^{er} et le 16^{ème} élément de la séquence restent invariants.

La figure 61 présente les 5 recompositions RGP utilisées dans les expériences. On rappelle également dans cette figure les permutations des PPP.

Pour chacun des 24 refrains, 11 variantes sont donc considérées :

- La version originale
- 5 recompositions PPP (également notées π)
- 5 recompositions RGP

Ce test perceptif est donc conduit sur 264 extraits musicaux distincts.

	$RGP_p(1)$	$RGP_p(2)$	$RGP_p(3)$	$RGP_p(4)$	$RGP_p(5)$	$RGP_p(6)$	$RGP_p(7)$	$RGP_p(8)$	$RGP_p(9)$	$RGP_p(10)$	$RGP_p(11)$	$RGP_p(12)$	$RGP_p(13)$	$RGP_p(14)$	$RGP_p(15)$	$RGP_p(16)$
RGP_1	1	4	5	3	8	2	7	6	9	14	11	10	12	15	13	16
RGP_2	1	2	9	10	3	4	11	12	5	6	13	14	7	8	15	16
RGP_3	1	2	5	6	3	4	7	8	9	10	13	14	11	12	15	16
RGP_4	1	5	9	13	2	6	10	14	3	7	11	15	4	8	12	16
RGP_5	1	3	9	11	2	4	10	12	5	7	13	15	6	8	14	16

	$\pi_p(1)$	$\pi_p(2)$	$\pi_p(3)$	$\pi_p(4)$	$\pi_p(5)$	$\pi_p(6)$	$\pi_p(7)$	$\pi_p(8)$	$\pi_p(9)$	$\pi_p(10)$	$\pi_p(11)$	$\pi_p(12)$	$\pi_p(13)$	$\pi_p(14)$	$\pi_p(15)$	$\pi_p(16)$
π_1	1	3	5	7	2	4	6	8	9	11	13	15	10	12	14	16
π_2	1	2	9	10	3	4	11	12	5	6	13	14	7	8	15	16
π_3	1	2	5	6	3	4	7	8	9	10	13	14	11	12	15	16
π_4	1	5	9	13	2	6	10	14	3	7	11	15	4	8	12	16
π_5	1	3	9	11	2	4	10	12	5	7	13	15	6	8	14	16

The figure displays a musical score for a piece in G major (one sharp) and 4/4 time. The score is organized into 11 staves. The first staff is labeled 'Originale'. The following five staves are labeled RGP_1 through RGP_5 , representing recomposed pieces. The final five staves are labeled π_1 through π_5 , representing another set of recomposed pieces. Each staff contains a single melodic line with various rhythmic patterns and accidentals.

Figure 61: Exemples de partitions de recompositions RGP et PPP

4.1.2 Protocole

Cette expérience perceptive a été menée sur une population de 66 personnes, âgées de 22 à 55 ans, et composée de 25% de femmes et de 75% d'hommes.

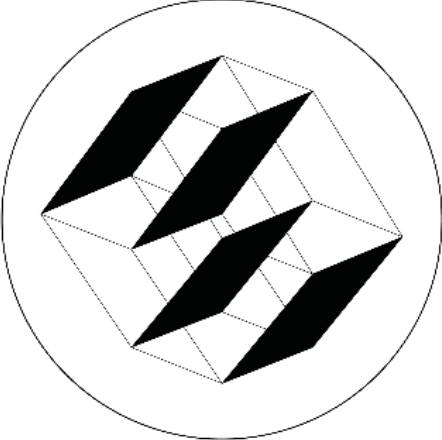
Afin d'éviter les biais de familiarité dus à la reconnaissance successive de tel ou tel extrait, on présente à chaque sujet une seule variante de chacun des 24 refrains (original ou recomposé) dans un ordre variable selon les sujets. Le test porte donc sur 24 items par sujet. Les variantes sont distribuées de façon le plus équilibrée possible pour chaque sujet, typiquement 2 (ou occasionnellement 3) du même type de permutation (original, 5 PPP et 5 RGP).

Après une brève présentation générale du test durant laquelle les sujets ne sont pas mis au courant des motivations scientifiques du test (voir texte figure 63), il leur est demandé de donner une réponse à deux questions formulées ainsi :

- Veuillez indiquer votre avis sur la cohérence de l'organisation musicale de cet extrait.
- Veuillez indiquer votre avis sur le niveau d'attrait musical de cet extrait.

Poser ces deux questions distinctement a pour but de décorréler au maximum la note qu'un sujet va attribuer à la cohérence musicale de l'extrait (question 1) par rapport à l'expression de ses goûts personnels pour cet extrait (question 2).

Les sujets répondent à ces questions en plaçant un curseur sur une échelle de 5 niveaux, d'évaluation (de « très mauvaise cohérence » à « très bonne cohérence » et de « très peu attrayant » à « très attrayant »). Les auditeurs ont également la possibilité de placer le curseur entre deux graduations, ce qui donne au total, 9 niveaux d'évaluation.



Merci d'avoir accepté de participer à ce test musical perceptif. Il vous prendra environ 15 minutes.

Vous allez entendre 24 passages musicaux au format midi et vous allez devoir leur attribuer deux appréciations :


- la première portant sur la cohérence de leur organisation (c'est-à-dire dans quelle mesure leur construction musicale vous paraît valide).
- la seconde sur l'attrait que vous leur trouvez (en d'autres termes, indiquer à quel point ce passage musical est à votre goût - ou pas).

Vous pourrez écouter le segment musical plusieurs fois si nécessaire avant de valider vos réponses.

L'analyse de l'ensemble des résultats obtenus sur le panel de participants permettra d'évaluer des travaux de recherche en cours.

Lorsque vous êtes prêts, indiquez votre nom ci-dessous puis validez.

Figure 62: Instructions données aux sujets participant aux tests perceptifs.



Extrait N° 1 / 24

Veillez indiquer votre avis sur la cohérence de l'organisation musicale de cet extrait
Please give your opinion as of the consistency of the musical construction of this passage

Très mauvaise cohérence Mauvaise cohérence Cohérence moyenne Bonne cohérence Très bonne cohérence

Très peu attrayant Peu attrayant Moyennement attrayant Attrayant Très attrayant

Figure 63: Echelle de notation des recompositions dans le cadre des tests perceptifs.

4.1.3 Évaluation des scores

Pour chaque sujet, on collecte donc 24 évaluations portant sur la cohérence et 24 évaluations sur l'attractivité correspondant à chacun des 24 extraits présentés.

On appelle r , une variable générique qui désigne une évaluation, elle peut être une évaluation de cohérence ou d'attractivité.

Les 24 évaluations r_{jk} produites par le sujet j ($1 \leq j \leq 66$, $1 \leq k \leq 24$) sont distribuées sur 11 variantes, O pour originale, π_i pour les recompositions PPP_i ($1 \leq i \leq 5$) et π_i^* pour les recompositions RGP_i ($1 \leq i \leq 5$).

On calcule les moyennes pour chaque type de variante ce qui nous donne 11 moyennes différentes pour un sujet j .

$$\mu_{Oj} = \frac{1}{m_{Oj}} \sum_O r_{jk} \text{ pour les refrains originaux.} \quad (4.1)$$

$$\mu_{ij} = \frac{1}{m_{ij}} \sum_{\pi_i} r_{jk} \text{ pour les variantes PPP.} \quad (4.2)$$

$$\mu_{ij}^* = \frac{1}{m_{ij}^*} \sum_{\pi_i^*} r_{jk} \text{ pour les variantes RGP.} \quad (4.3)$$

où m_{Oj} , m_{ij} et m_{ij}^* correspondent aux nombres de variantes originales et permutées testées par le sujet j , et sont égales à 2 ou 3, leur somme étant égale à 24.

Pour un sujet donné j , on calcule alors les *dégradations* δ_{ij} et δ_{ij}^* des scores de chaque variante de recompositions par rapport aux scores obtenus par les sections musicales dans leur forme originale.

$$\delta_{ij} = \mu_{Oj} - \mu_{ij} \text{ pour les variantes PPP.} \quad (4.5)$$

$$\delta_{ij}^* = \mu_{Oj} - \mu_{ij}^* \text{ pour les variantes RGP.} \quad (4.6)$$

On calcule finalement la quantité ρ_{ij} qui mesure, pour un sujet j , la préférence relative (positive ou négative) des recompositions PPP par rapport aux recompositions RGP.

$$\rho_{ij} = \mu_{ij} - \mu_{ij}^* = \delta_{ij} - \delta_{ij}^* \quad (4.7)$$

Finalement, on se concentre sur 3 scores :

$$\rho_j = \frac{1}{5} \sum_{i=1}^5 \rho_{ij} \quad (4.8)$$

$$\delta_i = \frac{1}{66} \sum_{j=1}^{66} \delta_{ij} \quad (4.9)$$

$$\delta_i^* = \frac{1}{66} \sum_{j=1}^{66} \delta_{ij}^* \quad (4.10)$$

ρ_j donne une estimation de la préférence globale (positive ou négative) des PPP par rapport aux RGP pour chaque sujet j .

δ_i (respectivement δ_i^*) mesure la dégradation du score obtenu par des recompositions PPP (respectivement RGP) en moyenne sur l'ensemble des sujets.

Ces quantités sont mesurées indépendamment pour les scores de cohérence et les scores d'attractivité.

4.1.4 Résultats

Les figures 64 et 65 représentent les distributions des scores de préférence en termes de cohérence et d'attractivité des recompositions PPP par rapport aux recompositions RGP. Les résultats des 66 sujets sont classés du sujet ayant la préférence la plus faible en faveur des PPP jusqu'à celui ayant la préférence la plus forte.

Dans les deux cas, les distributions sont clairement excentrées vers la partie positive du graphique, en particulier pour les scores de cohérence.

En effet, en ce qui concerne le score de cohérence, 57,5 sujets ont une préférence positive pour les recompositions PPP alors que 8,5 ont une préférence négative (87,1% contre 12,9%). On compte pour moitié dans les 2 catégories le sujet qui a une préférence nulle.

Le score moyen de jugement positif est de +0,89, alors que la moyenne des préférences négatives est seulement de -0.30.

Pour l'attractivité, les proportions sont toujours clairement en faveur des PPP, mais d'une façon un peu moins contrastée (74,2 % contre 25,8%) avec des scores moyens de +0.75 et de -0.55 respectivement.

D'après les figures 66 et 67 qui montrent la différence de score entre les recompositions PPP et les RGP homologues (c'est-à-dire les RGP qui induisent le même nombre de discontinuités et ont la même excursion), on remarque que la dégradation des scores est, dans les 2 cas, corrélée au nombre de discontinuités induites par la recomposition.

Globalement, ces résultats indiquent que les recompositions PPP ont tendance à être perçues comme moins perturbantes que les recompositions RGP par rapport au critère subjectif de cohérence musicale.

Ce résultat conforte donc l'hypothèse selon laquelle la préservation des systèmes d'implications contribue de façon importante au maintien de la cohérence de la structure d'une section musicale quant à son appréciation cognitive.



Figure 64: Résultats des tests perceptifs par sujet sur la question de la cohérence musicale.

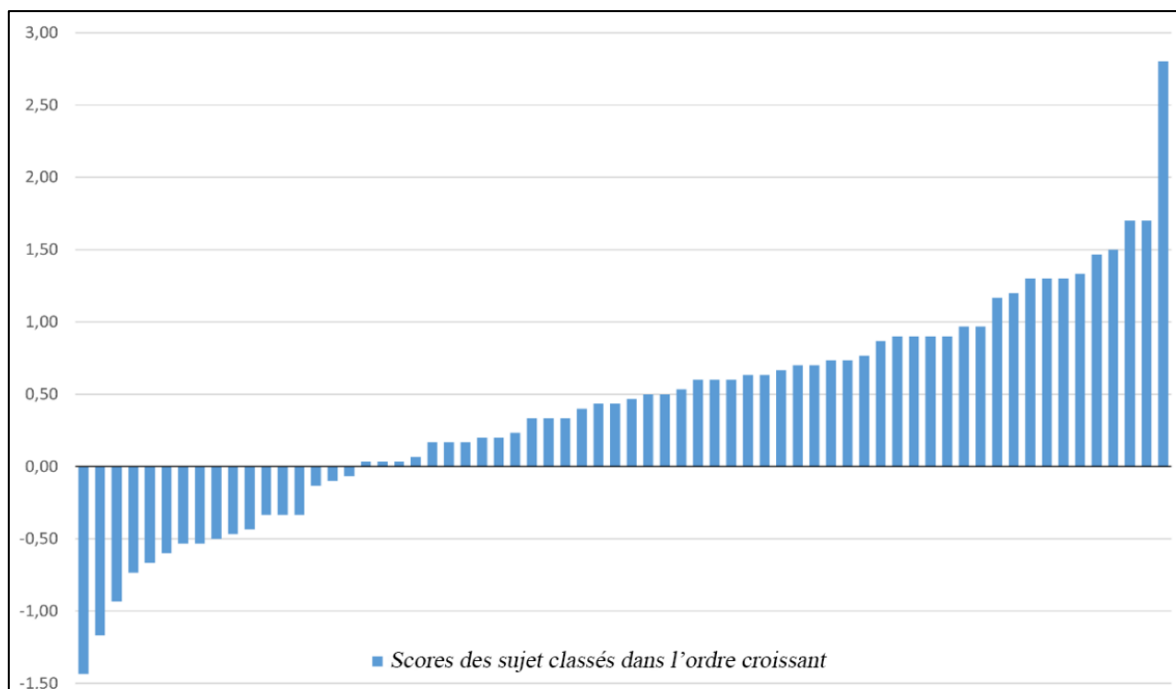


Figure 65: Résultats des tests perceptifs par sujet sur la question de l'esthétique musicale.

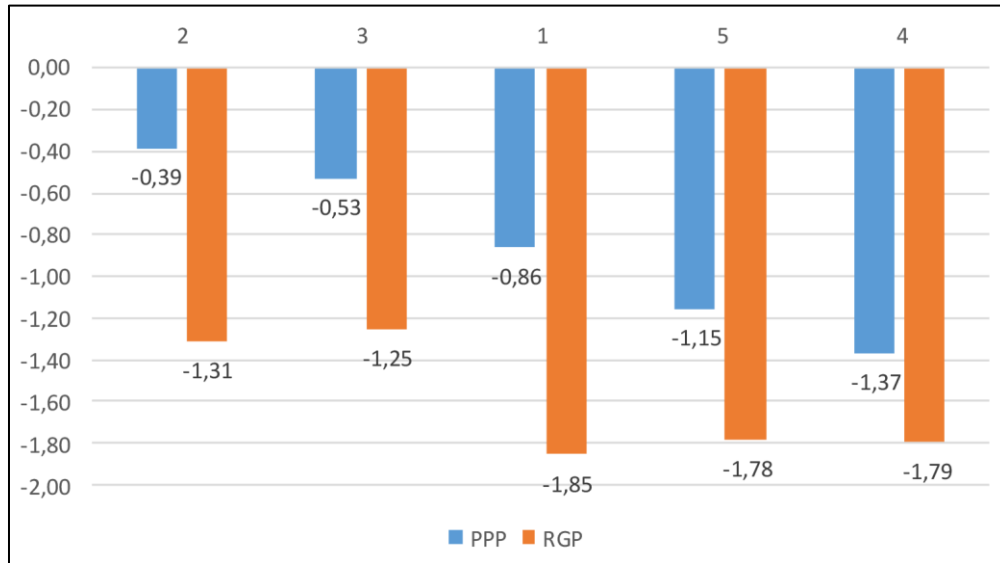


Figure 66: Comparaison des résultats des tests perceptifs en fonction du type de recomposition (critère de cohérence musicale).

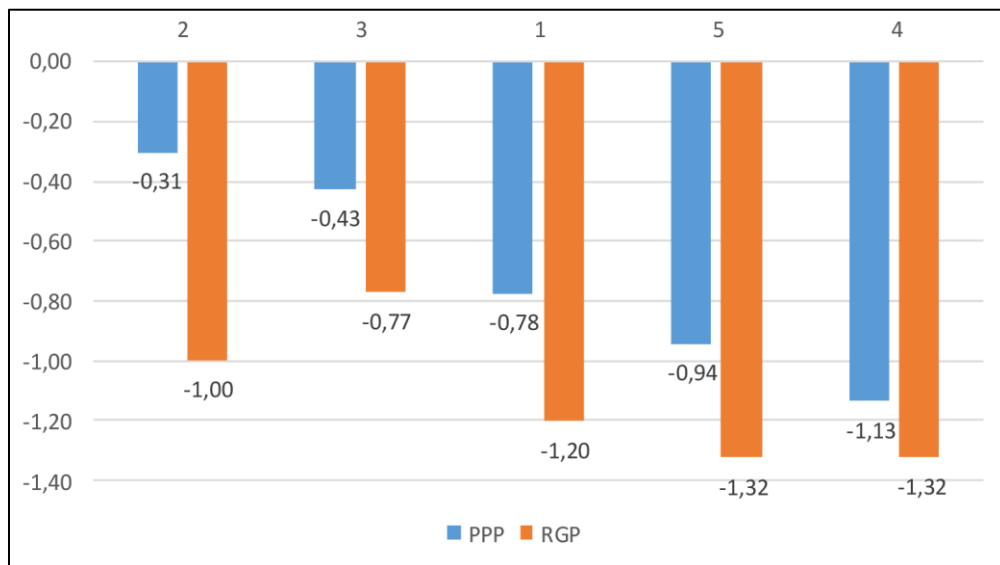


Figure 67: Comparaison des résultats des tests perceptifs en fonction du type de recomposition (critère de l'attractivité musicale).

4.2 Test 2 : évaluation perceptive des segmentations algorithmiques

Cette partie présente le second test perceptif mené dans ces travaux de thèse.

Après avoir montré avec le premier test que la conservation des systèmes d'implication avait un impact positif sur la perception musicale, on évalue maintenant l'appréciation perceptive des recompositions polytopiques générées à partir de différentes variantes d'algorithmes de segmentation.

4.2.1 Données

Dans cette seconde étude perceptive, les extraits sont issus du même corpus RWC POP, mais le test se déroule sur un nombre plus important d'extraits. On travaille ici sur un sous-ensemble du corpus constitué de 52 refrains.

Les mêmes types d'objets musicaux sont étudiés que dans le premier test : pour chaque extrait, on conserve son premier refrain avec une surface musicale de 8 mesures.

Toutefois, on ne conserve dans ce test que la ligne mélodique de chaque refrain avec une instrumentation de « piano » joué à leur tempo original.

On génère les 5 recompositions polytopiques variantes de l'ordre séquentiel de chacun de ces extraits, à partir de différentes segmentations déterminées algorithmiquement.

On compare les recompositions polytopiques entre elles, et sans introduire de variantes aléatoires.

Comme expliqué dans la section 3.3.3, un ensemble de 16 variantes algorithmiques sont étudiées dans ces travaux. Il n'est cependant pas possible de les tester toutes perceptivement car cela demanderait un nombre trop important de participants et/ou de tests par participants, ce qui serait très lourd à réaliser.

On décide donc de travailler sur un sous-ensemble de ces 16 variantes de méthodes et les sections suivantes détaillent comment ce sous-ensemble (de 6 variantes seulement) a été sélectionné.

4.2.2 Protocole de sélection des variantes de méthodes

Cette seconde étude perceptive présente certaines particularités par rapport à la première série de tests, dont il est nécessaire de tenir compte pour l'élaboration du protocole expérimental dans son ensemble.

D'après la section 3.3.3, on a sélectionné un ensemble de 16 méthodes algorithmiques d'estimations de décalage entre la surface musicale et la mesure. Néanmoins, il n'est pas envisageable de mener un test perceptif pour départager l'ensemble de ces 16 méthodes, pour des raisons d'étendue du test.

Pour cette raison, nous avons procédé à une sélection d'un sous-ensemble de méthodes qui sont retenues pour générer les extraits à tester.

Comme évoqué précédemment, plusieurs de ces méthodes donnent le même résultat de décalage θ et donc la même segmentation pour certains extraits. Ainsi, l'ensemble des recompositions polytopiques générées à partir de plusieurs méthodes d'estimations peuvent être partiellement identiques.

Ainsi, l'évaluation d'une recomposition polytopique donnée, un extrait donné et un décalage donné fournit simultanément des informations sur plusieurs variantes algorithmiques.

On s'appuie sur ce constat afin de recueillir une évaluation perceptive sur plusieurs méthodes algorithmiques à la fois, à partir de l'évaluation d'un seul extrait par un auditeur. Selon ce principe, on conçoit un test perceptif permettant d'obtenir un nombre satisfaisant de résultats sur un sous-ensemble de 6 méthodes algorithmiques, à partir d'un nombre raisonnable de participants.

Dans cette seconde série de test, il est présenté en moyenne 24 extraits à chaque auditeur, un chiffre comparable au test précédent mais qui peut varier légèrement d'un sujet à un autre.

Dans ces ~ 24 extraits, à peu près 1/4 (6 extraits) sont des refrains dans leur forme originale et le reste (environ 18 extraits) sont des recompositions polytopiques réparties équitablement selon les 6 variantes algorithmiques testées et les 5 PPP possibles (différentes de l'identité). Certains extraits contribuent à l'évaluation de plusieurs variantes algorithmiques ayant abouti à la même estimation de décalage optimal.

Ainsi, avec 166 participants x 24 tests environ (soient environ 4000 items testés), on obtient un nombre moyen de 1084 évaluations pour chacune des 6 méthodes d'estimations algorithmiques sélectionnées soient un peu plus de 6500 scores au total.

La section suivante est consacrée à la description du processus de sélection de ce sous-ensemble de 6 méthodes.

4.2.3 Sélections des méthodes algorithmiques testées

Méthodologie d'élimination successive

Comme discuté dans la section 3.3.3, l'estimation de la concordance entre variantes algorithmiques et annotations ne permet pas d'observer de tendance claire concernant l'appréciation des méthodes.

Dans cette section, on présente une méthodologie « d'élimination successive » permettant de définir une méthode algorithmique d'estimation de décalage fournissant un compromis entre les préférences des deux jeux d'annotations.

Cette méthodologie reprend le principe présenté dans le chapitre 3.3.2 :

- On calcule tout d'abord les scores de concordance de chacune des caractéristiques des méthodes.
- On choisit la caractéristique fournissant la meilleure concordance par rapport aux deux annotateurs (voir figure 57), ce qui correspond à la moitié des méthodes considérées et on « élimine » ainsi l'autre moitié des méthodes d'estimations étudiées.
- On réitère ce processus jusqu'à avoir déterminé toutes les caractéristiques successives.

Après avoir suivi ce principe sur les 16 méthodes que l'on cherche à départager, voici la méthode d'estimation Γ^* résultant de cette méthodologie :

$$\Gamma^* = \{\textit{Systèmes Indépendants}, \\ \textit{Primer comme antécédant à tous les motifs d'un système}, \\ \textit{Priorité au premier granule du motif}, \\ \textit{Propagation en cascade}, \\ \textit{Distance Musicale}\}$$

Sélection d'un sous ensemble de 6 méthodes d'estimation.

Le principe de la méthode par élimination successive permet de sélectionner une méthode qui effectue le meilleur compromis entre les deux jeux d'annotations.

On souhaite néanmoins sélectionner d'autres méthodes permettant de représenter les différentes variantes dans l'espace des 16 méthodes.

On choisit de tester un sous-ensemble de 6 d'entre elles qui présentent un intérêt de par leur degré de concordance aux annotations manuelles et/ou la diversité de leurs caractéristiques.

Les 6 méthodes algorithmiques qui sont donc à l'étude dans ce test sont :

- Les deux méthodes les plus concordantes respectivement avec les deux jeux d'annotations manuelles A_1 et A_2 : elles sont notées Γ_I et Γ_{II}
- La méthode déterminée par la méthodologie « d'élimination successive » : Γ_{III} .
- La méthode la moins concordante avec les annotations, nommée Γ_{IV} .
- Deux autres méthodes permettant de tester des caractéristiques qui équilibrent celles sélectionnées dans les 4 premières méthodes. Elles sont nommées Γ_V et Γ_{VI} .

Au total, chaque caractéristique est présente 3 fois sous une forme et 3 fois sous une autre, dans la liste considérée (voir Figure 68).

	Γ_I	Γ_{II}	Γ_{III}	Γ_{IV}	Γ_V	Γ_{VI}
<i>Complexité entre systèmes</i>	Indépendants	Super-Système	Indépendants	Indépendants	Super-Système	Super-Système
<i>Complexité interne à un système</i>	Seulement «Primer»	Seulement «Primer»	Seulement «Primer»	Recherche min. «Contraste»	Recherche min. «Contraste»	Recherche min. «Contraste»
<i>Propagation du transport</i>	1 ^{er} intervalle	Cascade	Cascade	1 ^{er} intervalle	1 ^{er} intervalle	Cascade
<i>Point d'ancrage des granules</i>	Recherche bilatérale	Recherche bilatérale	Standard	Standard	Recherche bilatérale	Standard
<i>Distance élémentaire entre granules</i>	Musicale	Musicale	Musicale	Musicale	Musicale	Musicale

Figure 68 : Récapitulatif des méthodes évaluées dans le second test perceptif.

Comme discuté dans la section 4.2.2, Il arrive que plusieurs des méthodes d'estimations produisent le même décalage θ , et donc la même segmentation. Ainsi les 5 recompositions PPP générées à partir de ce décalage seront identiques.

Cela permet donc d'obtenir une appréciation sur plusieurs méthodes à la fois, à partir de l'évaluation d'un seul extrait :

Par exemple, pour l'extrait n°15, les 6 méthodes donnent le même décalage. Même si, dans ce cas, l'évaluation ne permet pas de départager les méthodes algorithmiques, ce résultat est intéressant pour comparer le score d'une recomposition à celui d'un extrait dans sa forme originale.

En revanche pour l'extrait n°10, seulement 2 des 6 méthodes donnent un décalage identique. En moyenne sur chaque extrait, on dénombre typiquement entre 2 et 3 décalages différents définis par ces 6 méthodes algorithmiques.

Dans le cadre de cette seconde étude perceptive, on évalue perceptivement 52 refrains originaux et 690 recompositions polytopiques.

4.2.4 Protocole du test perceptif

Du point de vue du sujet testé, le protocole de cette seconde étude perceptive est très similaire à celui du premier. La même plateforme web est utilisée pour collecter les résultats de chaque participant.

Le sujet utilise donc la même interface de collecte des réponses que dans le test présenté précédemment : le sujet doit évaluer la cohérence et l'attrait de chaque extrait à l'aide de deux curseurs donnant chacun 9 niveaux d'évaluation, allant de très « mauvaise cohérence/très peu attrayant » à « très bonne cohérence/très attrayant » (voir figure 63).

Ce second test est conduit sur un panel de 166 participants recruté par diffusion d'appel à participation sur les listes de contacts de l'auteur de cette thèse et de son directeur de thèse.

Certaines personnes pouvant être réticentes à l'idée que des informations personnelles les concernant soient stockées dans le cadre d'un tel test, aucune information d'âge, de sexe ou de culture musicale n'a été demandée pour ce test, de sorte à obtenir le plus de réponses possibles.

Il est donc impossible d'estimer le profil de ce panel, qu'il s'agisse des tranches d'âge, de la proportion d'hommes et de femmes, de leur type de culture musicale, etc...

Chaque auditeur évalue un nombre moyen de 24 mélodies (8 mesures), et attribue à chaque extrait, une note de cohérence et une note d'attractivité.

Parmi l'ensemble des extraits présentés à un auditeur, une proportion d'1/4 de ces extraits sont des refrains dans leur forme originale. Les autres extraits présentés sont des recompositions polytopiques réparties équitablement selon la variante de la fonction PPP utilisée.

De plus, et de la même façon que pour le premier test, afin d'éviter les biais d'accoutumances, chaque auditeur n'entend qu'une seule version (recomposée ou non) de chaque refrain qui lui est présentée, mais comme il y a maintenant 52 refrains dans le corpus de test, un auditeur donné ne les entend pas tous.

4.2.5 Évaluation des scores

De sorte à être cohérent avec les méthodes d'évaluations des scores du premier test perceptif, nous utiliserons dans cette section les mêmes principes pour évaluer les scores collectés.

Pour chaque auditeur, on collecte N_j évaluations portant sur la cohérence et N_j évaluations sur l'attractivité (N_j vaut en moyenne 24).

On appelle r , une variable générique qui désigne une évaluation, elle peut être une évaluation de cohérence ou d'attractivité.

Les N_j évaluations r_{jk} produits par le sujet j ($1 \leq j \leq 166$, $1 \leq k \leq N_j$) sont distribuées sur 7 variantes, O pour originale et les Γ_i correspondant à des recompositions polytopiques issues des 6 méthodes algorithmiques d'estimations ($1 \leq i \leq 6$).

On calcule ensuite les moyennes pour chaque type de variante et on obtient les 7 moyennes suivantes :

$$\begin{aligned}\mu_{0j} &= \frac{1}{m_{0j}} \sum_{O} r_{jk} , \text{ pour les refrains originaux.} \\ \mu_{ij} &= \frac{1}{m_{ij}} \sum_{\Gamma_i} r_{jk}\end{aligned}\quad (4.12)$$

pour les 6 méthodes algorithmiques (toutes PPP confondues).

Le dénominateur m_{0j} correspond au nombre de refrains originaux et m_{ij} au nombre de variantes par méthodes algorithmiques présentées à l'auditeur j . Leur somme est égale au nombre total N_j d'évaluations pour cet auditeur.

Comme pour le test précédent, afin de comparer les résultats de chaque auditeur sur une base commune, on normalise les notes attribuées en fonction de l'appréciation relative des extraits recomposés par rapport aux extraits dans leur forme originale.

On définit ainsi la variable δ_{ij} rendant compte de la différence d'appréciation entre les refrains originaux et les recompositions générées à partir d'une méthode algorithmique pour l'auditeur j et pour chacune des i méthodes algorithmiques.

$$\delta_{ij} = \mu_{ij} - \mu_{0j} \quad (4.13)$$

Cette variable indique donc le niveau de dégradation des notes attribuées aux recompositions polytopiques générées par segmentation algorithmique par rapport à la note moyenne des refrains dans leur forme originale, pour un auditeur donné.

On calcule finalement la moyenne obtenue pour chaque méthode algorithmique i sur l'ensemble des 166 sujets ayant participé au test :

$$\delta_i = \frac{1}{166} \sum_{j=1}^{166} \delta_{i,j} \quad (4.14)$$

Le score δ_i mesure la dégradation du score obtenu par les recompositions engendrées par chaque méthode algorithmique en moyenne sur l'ensemble des auditeurs. Ces quantités sont mesurées indépendamment pour les scores de cohérence et les scores d'attractivité.

4.2.6 Résultats

D'après les résultats du premier test perceptif, on a vu que les recompositions polytopiques conservant la logique des systèmes d'implications sont évaluées perceptivement comme musicalement « plus acceptables » que des recompositions aléatoires.

Dans le cadre de ce second test, on peut visualiser sur les figures 66, 67, 70 et 72 la dispersion des scores de dégradation pour les différentes méthodes et constater qu'elle est plus faible que dans le premier test.

Ceci s'explique probablement par le fait que, dans ce second test, toutes les recompositions sont des recompositions polytopiques dont la valeur des scores est plus concentrée que dans le cas de tests mélangeant des recompositions polytopiques et des recompositions aléatoires, comme c'était le cas du 1^{er} test.

On note également que la moyenne des scores de dégradation des recompositions (toutes polytopiques) est moindre que celle des recompositions polytopiques du premier test. Cela peut s'expliquer par le fait que dans le premier test, on avait conservé la mélodie, la rythmique, la basse et l'accompagnement harmonique, de ce fait les recompositions polytopiques étant appliquées sur plus d'éléments, plus de perturbations étaient ainsi induites.

Enfin, compte tenu des faibles écarts de scores enregistrés entre les différentes méthodes dans cette nouvelle série de tests, il est évident que les résultats obtenus ne permettent pas de tirer de conclusions définitives. Mais nous commentons ci-après, les tendances observées avec implicitement toutes les précautions d'usages quant au caractère significatif des résultats.

Score de Cohérence Musicale

On s'intéresse tout d'abord aux résultats concernant la cohérence musicale.

Ce critère perceptif a pour but d'évaluer l'acceptabilité de la « construction musicale » des différentes recompositions polytopiques.

Comme expliqué dans la section 4.2.5, les scores de cohérence des recompositions polytopiques sont comparés aux scores obtenus par les extraits dans leur forme originale, on peut donc interpréter ces résultats comme une mesure de dégradation moyenne.

D'après les figures ci-dessous, on voit que la méthode déterminée par *élimination successive* Γ_{III} (cf. section 4.2.3) obtient le meilleur score avec une dégradation moyenne de 0,55 par rapport aux scores des extraits originaux.

Les méthodes les plus concordantes avec le premier jeu d'annotation Γ_I et avec le second Γ_{II} sont les 2^{ème} et 3^{ème} meilleures méthodes avec un score de 0,57 et 0,59 graduation par rapport aux extraits originaux.

Les 3 méthodes restantes obtiennent des scores de dégradation plus élevés : 0,63 pour la méthode Γ_{IV} , 0,64 pour la méthode Γ_{VI} et 0,72 pour Γ_V .

En dépit du relativement faible écart de score entre les différentes méthodes (dont on ne sait pas interpréter la significativité), on observe que les résultats de l'évaluation perceptuelle des méthodes d'estimations suivent globalement la même tendance que les jeux d'annotations de référence.

	Γ_{III}	Γ_I	Γ_{II}	Γ_{IV}	Γ_{VI}	Γ_V
<i>Complexité entre systèmes</i>	Indépendants	Indépendants	Super-Système	Indépendants	Super-Système	Super-Système
<i>Complexité interne à un système</i>	Seulement «Primer»	Seulement «Primer»	Seulement «Primer»	Recherche min. «Contraste»	Recherche min. «Contraste»	Recherche min. «Contraste»
<i>Propagation du transport</i>	Cascade	1 ^{er} intervalle	Cascade	1 ^{er} intervalle	Cascade	1 ^{er} intervalle
<i>Point d'ancrage des granules</i>	Standard	Recherche bilatérale	Recherche bilatérale	Standard	Standard	Recherche bilatérale
<i>Distance élémentaire entre granules</i>	Musicale	Musicale	Musicale	Musicale	Musicale	Musicale
Cohérence	-0,55	-0,57	-0,59	-0,63	-0,64	-0,72

Figure 69: Tableau présentant les méthodes algorithmiques et leurs scores de cohérence (Second test perceptif)

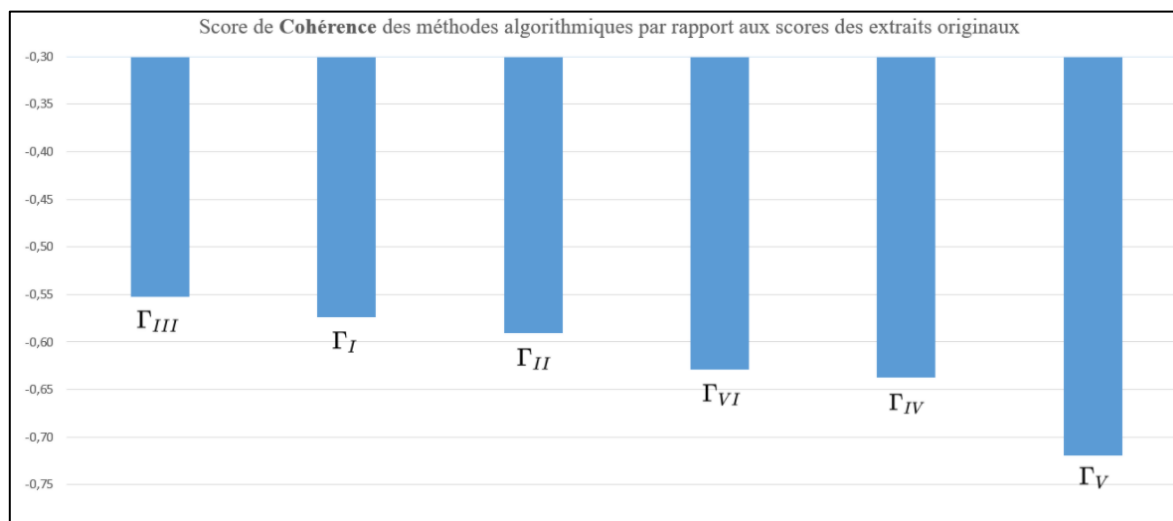


Figure 70: Résultats des scores de cohérence du second test perceptif

Attractivité

Comme expliqué dans la section précédente, l'estimation de score d'attractivité permet aux auditeurs d'évaluer séparément l'aspect structurel et l'aspect esthétique des extraits.

Globalement, on remarque que les scores d'attractivité des méthodes algorithmique sont moins dégradés par recomposition, par rapport aux extraits originaux, que dans le cas de l'évaluation de la cohérence.

De plus, on remarque que la méthode Γ_{III} déterminée par la méthodologie « d'élimination successive » présente ici encore les meilleurs scores par rapport aux autres méthodes, mais d'une courte tête par rapport à la méthode Γ_{IV} , avec, juste après, les méthodes Γ_I et Γ_{II} .

L'ordre de préférence des méthodes n'est pas le même que pour le jugement de cohérence musical, et la méthode arrivant en second est pourtant celle qui était la moins corrélée aux annotations de référence...

Est-ce dû à des particularités algorithmiques de la méthode qui renforce l'attrait musical des extraits, ou simplement à des fluctuations statistiques non significatives ? Il est difficile de conclure à ce stade et des investigations complémentaires seraient nécessaires pour mieux comprendre ce résultat.

Mais peut-être, en effet, l'attractivité d'un extrait musical pour certains auditeurs n'est pas intégralement liée à la cohérence de sa structure (surtout pour des morceaux tels que ceux de RWC Pop, qui sont très particuliers).

	Γ_{III}	Γ_{IV}	Γ_{II}	Γ_I	Γ_{VI}	Γ_V
<i>Complexité entre systèmes</i>	Indépendants	Indépendants	Super-Système	Indépendants	Super-Système	Super-Système
<i>Complexité interne à un système</i>	Seulement «Primer»	Recherche min. «Contraste»	Seulement «Primer»	Seulement «Primer»	Recherche min. «Contraste»	Recherche min. «Contraste»
<i>Propagation du transport</i>	Cascade	1 ^{er} intervalle	Cascade	1 ^{er} intervalle	Cascade	1 ^{er} intervalle
<i>Point d'ancrage des granules</i>	Standard	Standard	Recherche bilatérale	Recherche bilatérale	Standard	Recherche bilatérale
<i>Distance élémentaire entre granules</i>	Musicale	Musicale	Musicale	Musicale	Musicale	Musicale
Attrait	-0,35	-0,36	-0,39	-0,42	-0,44	-0,54

Figure 71: Tableau présentant les méthodes algorithmiques et leurs scores d'attrait (Second test perceptif)

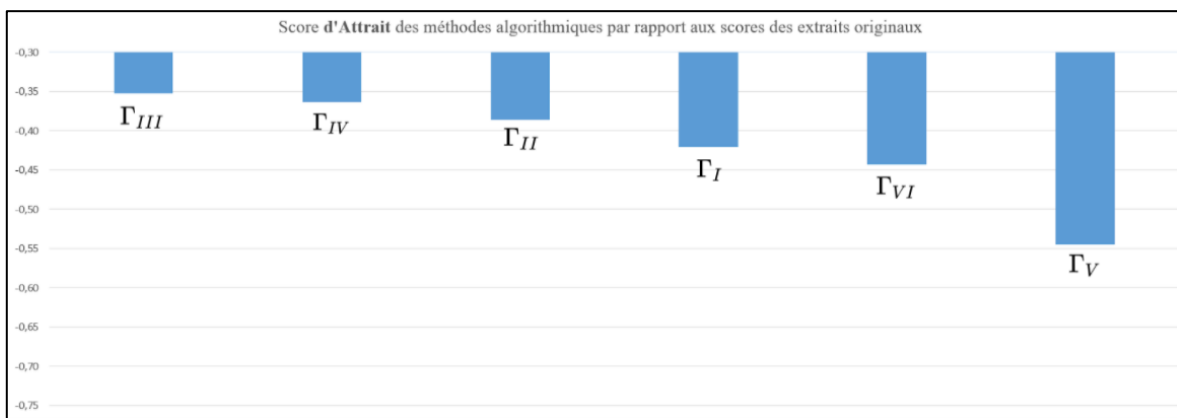


Figure 72: Résultats des scores d'attractivité du second test perceptif

4.3 Conclusion test perceptif

Les résultats obtenus dans le cadre du 1^{er} test présenté dans ce chapitre suggèrent assez nettement que les recompositions polytopiques suscitent une nette préférence par rapport à des recompositions qui ne suivent pas cette logique, tout au moins sur la base de données utilisée pour ces tests.

La seconde étude analyse la préférence relative entre des recompositions polytopiques issues de segmentations de plusieurs méthodes algorithmiques.

Cette seconde série de résultats obtenus sont plus nuancés et il est difficile d'affirmer qu'une méthode algorithmique est nettement plus performante que les autres, même si la méthode algorithmique obtenue par élimination successive arrive en tête.

Le fait que les résultats suivent globalement la tendance des scores de concordance entre les résultats des méthodes algorithmiques et des deux jeux d'annotations manuelles est une indication intéressante.

On peut souligner également que les scores obtenus par les différentes recompositions polytopiques ne sont pas très dégradés par rapport aux scores obtenus par les sections musicales dans leur forme originale ce qui suggère que l'invariance polytopique est prédominante devant l'estimation précise d'un décalage (supposément) optimal de la surface musicale.

Globalement, ces études perceptives permettent de montrer que les recompositions polytopiques sont capables de générer des contenus musicaux intéressantes perceptivement.

Le chapitre suivant présente comment ces concepts ont été utilisés afin de développer une application mobile visant à permettre à un large public de s'emparer des processus décrits dans ces travaux à des fins créatives.

Chapitre 5

Interface de Création Musicale Par Recomposition Polytopique

Cette partie présente la réalisation d'une interface créative permettant la simulation des concepts développés dans cette thèse afin de générer des variations de contenus par recompositions polytopiques.

5.1 Interfaces de variations existantes

Il existe plusieurs interfaces visant à faciliter la création de variations de contenu musical, c'est le cas par exemple de l'application web Unmixer développé dans le cadre des travaux de Jordan B. L. Smith (Smith J. B., 2019) et s'appuyant sur des techniques de séparation de sources pour isoler des boucles intéressantes à agencer différemment.

Il existe également des systèmes facilitant l'improvisation homme/machine comme l'application Improtek développée par Jérôme Nika (Nika, 2017) qui s'appuie sur un modèle musical génératif guidé permettant le réagencement de passages musicaux dans le cadre d'un dialogue logiciel/musicien.

L'application présentée ici permet de générer des variations polytopiques, à partir d'une section musicale segmentée en « re-séquençant » son contenu selon l'ordre guidé par la logique des différents systèmes d'implication de la séquence originale.

5.2 Application Tessy

Dans l'optique de « démocratiser » les concepts de variations polytopiques présentés dans ces travaux, nous avons collaboré avec l'entreprise *Kercodex* afin de développer une application utilisable sur tablette et/ou téléphone mobile qui soit accessible à un large public. Ces travaux de valorisation ont débouché sur une version prototype d'application que nous avons provisoirement dénommée *Tessy* (un diminutif du mot Tesseract !).

L'application permet à un utilisateur « naïf » de manipuler des sections musicales dans une représentation non séquentielle et d'en générer un grand nombre de variations (polytopiques ou non).

Ces variations sont guidées pour exploiter les propriétés structurelles présentées dans ces travaux, mais l'utilisateur est également libre de s'en affranchir.

L'interface de Tessy comporte 16 boutons tactiles interactifs correspondant aux 16 motifs structurants d'une section musicale prédéfinie.

Cette représentation vise à ce que l'utilisateur s'approprié la logique non séquentielle d'une section musicale, puisse en faire l'expérience, et développer sa compréhension intuitive des concepts de délinéarisation de contenus musicaux.



Figure 73: Présentation de l'interface de l'application Tessy

5.2.1 Modes de fonctionnement

Une section est représentée sous forme de 16 boutons qui correspondent à chacun des 16 motifs de la section et différents modes de fonctionnement sont implémentés.

En mode « lecture par segment » l'utilisateur peut activer successivement chacun des 16 boutons pour faire jouer les motifs dans un ordre quelconque.

Il peut alors réorganiser le séquençement des motifs du contenu musical et improviser une nouvelle séquence dans l'ordre de son choix.



Figure 74: Illustration de l'application Tessy en mode lecture par segment.

Ce mode par défaut de l'application permet de séquencer les motifs en déterminant leur ordre de lecture par pression successive sur chacun des boutons.

La séquence commence à être jouée lorsque l'élément 1 est actionné (toujours en premier, donc), et l'utilisateur est ensuite libre de définir l'ordre dans lequel les prochains motifs seront joués.

Toutefois, afin de guider l'utilisateur et de lui indiquer comment préserver au mieux, à un moment donné, les systèmes d'implications durant la définition de sa séquence, un système de colorisation des boutons indique la logique 'optimale' induite par les systèmes d'implications déjà réalisés et donc programme les boutons les mieux adaptés à cette logique.

On peut comprendre ce principe comme une forme de complétion automatique de la séquence amorcée par l'utilisateur pour conserver un certain niveau de logique systémique pour la suite de la séquence.



Figure 75: Présentation de la logique de "complétion automatique" pour conservation de la logique systémique dans Tessy

Lorsque la séquence est lancée, un code couleur est utilisé, comme cela est illustré sur la figure 75.

Les boutons colorés en vert correspondent aux motifs déjà joués, le bouton en rouge est le motif qui est en train d'être lu, les boutons colorés en jaunes indiquent le motif qui permet de conserver la logique systémique et les boutons violets sont les motifs qui restent à jouer et dont on peut programmer ensuite le séquençage.

5.2.2 Variations polytopiques

En plus des modes de jeu « libres » présentés dans la section précédente, il est possible de définir directement l'ordre des motifs de la séquence grâce aux boutons numérotés en bas de l'écran.

Par défaut, ces boutons définissent un ordre de lecture des motifs qui suit l'une des 6 fonctions de recompositions polytopiques PPP présentées dans ces travaux.

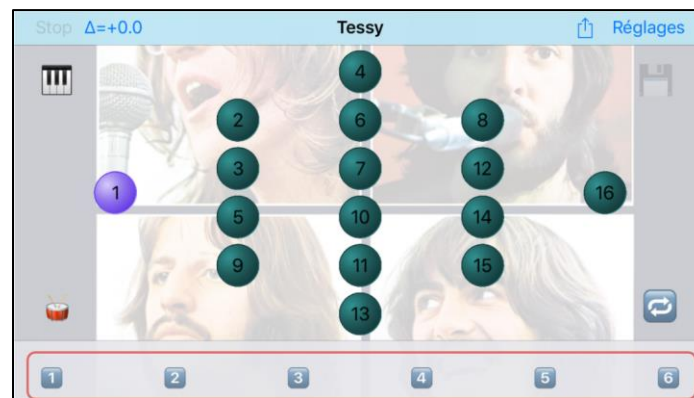


Figure 76: Interface de contrôle permettant de lancer une séquence PPP dans l'application Tessy.

Cependant, il est également possible pour l'utilisateur d'utiliser ces boutons numérotés comme autant de *cases mémoires* lui permettant de stocker les séquences de son choix.

Pour cela il suffit à l'utilisateur d'appuyer rapidement 2 fois sur un de ces numéros pour déclencher un menu lui permettant d'enregistrer la dernière séquence jouée dans l'application.



Figure 77: Présentation des cases mémoires dans l'application Tessy.

5.2.3 Visualisation temporelle

La prise en main de la vue délinéarisée n'est pas quelque chose d'immédiatement intuitif pour un utilisateur non familier du principe.

Afin de mettre en correspondance la vue temporelle séquentielle et la vue polytopique délinéarisée utilisée dans cette application, une jauge de progression temporelle est présente en bas de l'écran.

Deux modes d'affichages réglables sont disponibles dans les options.



Figure 78: Présentation des deux modes de visualisation temporelle de l'application Tessy sur la même recomposition PPP.

Le premier mode (à gauche) permet la visualisation séquentielle progressive de la jauge temporelle et les indices des motifs structurants sont permutés selon l'ordonnancement défini par la séquence polytopique.

Le deuxième mode (à droite) permet également de visualiser l'évolution de la lecture de la séquence, mais ici les indices temporels des motifs ne sont pas permutés.

En revanche la progression temporelle ne s'effectue pas de façon séquentielle : chaque zone temporelle attribuée à l'un des motifs est remplie au moment où celui-ci est joué même s'il n'est pas juxtaposé séquentiellement au motif actif.

Ce dernier mode permet de mieux appréhender la délinéarisation de la séquence.

Une fois la recomposition effectuée, l'utilisateur a la possibilité de sauvegarder sa « performance » au format MP3 et l'exporter pour diffusion.

Dans ce prototype, l'application propose plusieurs thèmes musicaux connus qu'il est possible de manipuler. Dans une future version, il est prévu que l'utilisateur puisse importer lui-même ses propres compositions et utiliser l'application pour en générer de multiples variations.

Si l'utilisateur souhaite avoir un contrôle avancé sur la production de sa variation polytopique, plusieurs options sont modifiables via le menu « réglages ».

Il est possible de changer l'instrumentation du thème musical et choisir entre plusieurs synthétiseurs, basses, batteries et piano.

L'utilisateur a également la possibilité de régler les différents niveaux sonores de chaque instrument et donc de mixer différemment la séquence.

Il est enfin possible de changer le décalage de la surface musicale avec la première barre de mesure, permettant ainsi de définir de nouveaux motifs en décalant l'origine des temps en fonction de la valeur de ce paramètre. Le paramètre de décalage est ici représenté par le symbole Δ (qui correspond à θ dans les sections précédentes).

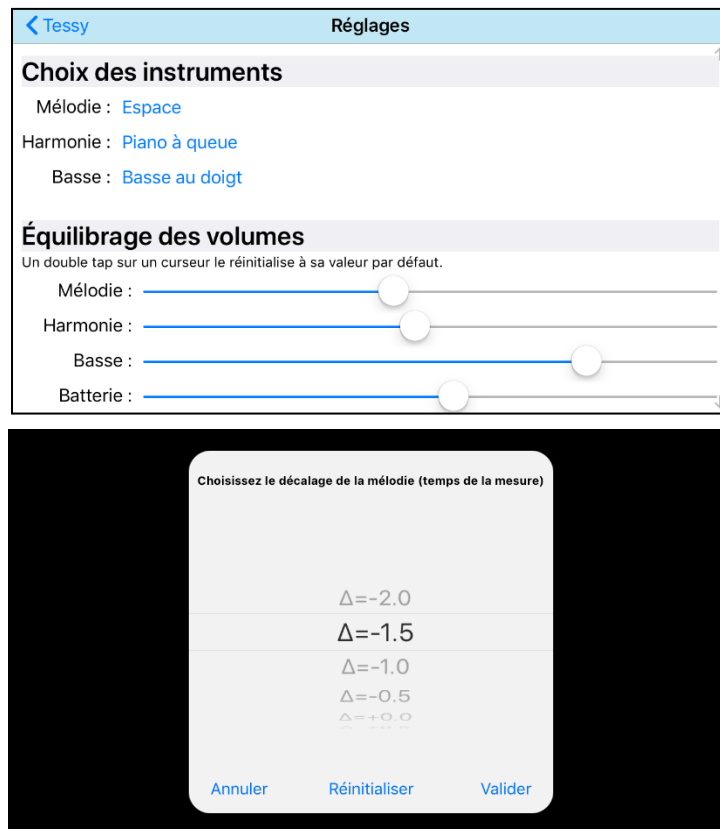


Figure 79: Réglages avancés de l'application Tessy

5.3 Perspectives de l'application Tessy

L'application Tessy en est encore au stade de prototype, et plusieurs améliorations sont en cours de développement.

Il est notamment prévu dans une prochaine version de faire en sorte que l'utilisateur puisse charger une séquence musicale de son choix, à partir d'un catalogue ou une composition qu'il aura effectuée préalablement. On prévoit également une refonte du design afin de rendre l'environnement plus attractif et ergonomique.

Une fois ces améliorations effectuées, l'application sera mise à disposition sur l'Apple Store d'applications et disponible pour un large public.

Une autre perspective envisagée en vue de la diffusion de ces travaux à plus long terme, serait de développer une solution permettant l'utilisation des principes présentés dans cette application dans un logiciel séquenceur (DAW). On pense notamment à un outil Max permettant ce type de manipulation sur des clips audio ou MIDI.

Chapitre 6

Conclusions et débouchés

Basés sur des travaux qui se développent depuis plusieurs années à l'IRISA autour du modèle « Système et Contraste », la présente thèse s'est intéressée à l'exploration, la validation et l'utilisation de ce modèle de structure musicale multi-échelle pour la description et la manipulation de séquences musicales et plus particulièrement mélodiques.

Cette thèse vient compléter un certain nombre de résultats concernant le modèle polytopique d'organisation de l'information mélodique sous trois angles : de nouveaux algorithmes de décomposition, une batterie de tests perceptifs et un prototype d'outil créatif pour la recomposition musicale.

Le principe de recomposition polytopique au cœur de cette thèse consiste à créer un nouveau contenu à partir de réorganisations de l'ordre temporel des motifs musicaux structurants d'un contenu d'origine, tout en préservant l'invariance de sa description sur un support polytopique, celle-ci rendant compte des similarités entre motifs à différentes échelles de temps. Cela revient en quelque sorte à intervertir la hiérarchie des échelles de temps qui régissent les relations entre motifs de la séquence mélodique.

Pour évaluer la pertinence de cette approche, une première série de tests perceptifs a été effectuée (Gillot, 2019) en faisant évaluer à un panel d'auditeurs la cohérence musicale de recompositions polytopiques par rapport à des recompositions ordonnancées aléatoirement, qui ne respectent pas l'invariance par permutation d'échelles.

Les résultats de ces tests étayent la pertinence de la recomposition polytopique, tout en mettant en évidence une influence du nombre et de l'étendue des discontinuités subis par le contenu d'origine après recomposition. Réciproquement, les résultats obtenus tendent à accréditer les hypothèses sous-jacentes au modèle polytopique comme significatives pour rendre compte de la structure musicale.

La contribution algorithmique de ce travail porte sur de nouveaux procédés de décomposition permettant de formaliser puis d'optimiser la décomposition d'une séquence mélodique en motifs. Ceci s'effectue sur la base de la minimisation d'un coût de description d'une section musicale qui peut s'interpréter comme une estimation de sa complexité algorithmique conditionnellement à un décalage constant de sa surface musicale. Ce coût se calcule en termes de transport minimal entre motifs, et différentes variantes ont été définies et développées dans cette thèse.

Une méthodologie en trois étapes est ensuite proposée pour évaluer et comparer les différentes variantes algorithmiques en l'absence d'une segmentation de référence objective des mélodies en motifs. On s'appuie alors dans un premier temps sur des segmentations manuelles produites sur la base de critères subjectifs. Puis, dans un second temps, on évalue les différentes variantes algorithmiques en fonction de leur concordance avec les segmentations, ce qui permet de retenir un sous-ensemble de variantes plus ou moins corrélées avec les segmentations manuelles. Enfin, dans une troisième et dernière étape, on

gènere et on teste auprès d'un panel d'auditeurs humains l'acceptabilité musicale des recompositions polytopiques avec ces différentes variantes.

Les résultats obtenus selon ce protocole permettent de classer les différentes variantes d'algorithmes, sans pour autant présenter des écarts très nets. Ceci indique qu'il existe une préférence faible entre critères de segmentation, sans pour autant montrer de tendance décisive. Toutefois, cette étude met incidemment en évidence que, sur le corpus considéré, la similarité entre éléments mélodiques doit préférablement s'opérer dans le domaine harmonique que dans le domaine chromatique.

Une troisième contribution de cette thèse est d'avoir permis la mise en place d'outils d'aide à la création musicale exploitant le principe des recompositions polytopiques pour manipuler un contenu musical dans des dimensions jamais explorées jusqu'à présent, à notre connaissance.

Il s'agit en fait de procéder à une transformation isométrique du contenu musical dans l'espace « mésoscopique » de ses échelles de temps qui permet de rapprocher des relations « lointaines » et d'éloigner des similarités immédiates, avec un effet tout à fait intéressant en termes d'expérience musicale et de potentiel créatif.

Une preuve de concept a été développée dans la cadre de cette thèse en collaboration avec l'entreprise Kercodex sous forme d'une application mobile permettant l'utilisation des principes du modèle Système & Contraste et des recompositions polytopiques sur des morceaux (pour l'instant) en format MIDI.

Cette interface permet de générer interactivement des variations polytopiques d'un contenu musical d'origine. Elle permet tout à la fois d'illustrer les concepts fondamentaux à l'origine de cette thèse, mais aussi les utiliser de façon créative et/ou ludique auprès d'un large public. A cet égard, même s'il reste un potentiel certain d'amélioration algorithmique des méthodes abordées dans cette thèse, les dimensions pédagogiques, exploratoires et inventives de ce travail constituent des retombées applicatives à approfondir.

Bibliographie

- Austin, W. W. (1966). *Music in the 20th Century: from Debussy through Stravinsky*. WW Norton.
- Bigo, L., Spicher, A., Ghisi, D., Andreatta, M. (2014, September). Spatial transformations in simplicial chord spaces. In *Proceedings ICMC| SMC| 2014* (pp. 1112-1119).
- Bimbot, F., Deruty, E., Sargent, G., Vincent, E. (2012, October). Semiotic structure labeling of music pieces: Concepts, methods and annotation conventions. In *13th International Society for Music Information Retrieval Conference (ISMIR)*.
- Bimbot, F., Deruty, E., Sargent, G., Vincent, E. (2016). System & contrast: a polymorphous model of the inner organization of structural segments within music pieces. *Music Perception: An Interdisciplinary Journal*, 33(5), 631-661.
- Côté-Lapointe, S. (2013). *Approches d'analyse et descripteurs dans l'indexation et la classification de la musique*.
- Cuthbert, M. S., Ariza, C. (2010). *music21: A toolkit for computer-aided musicology and symbolic music data*.
- Deruty, E., Bimbot, F., Van Wymeersch, B. (2013). *Methodological and musicological investigation of the System & Contrast model for musical form description (Doctoral dissertation, INRIA)*.
- Durand, B., Zvonkin, A. (2007). Kolmogorov complexity. In *Kolmogorov's heritage in mathematics* (pp. 281-299). Springer, Berlin, Heidelberg.
- Gillot, V., Bimbot, F. (2019, May). Polytopic reconfiguration: a graph-based scheme for the multiscale transformation of music segments and its perceptual assessment. In *SMC 2019-16th Sound & Music Computing Conference* (pp. 1-8).
- Goto, M., Hashiguchi, H., Nishimura, T., Oka, R. (2002, October). RWC Music Database: Popular, Classical and Jazz Music Databases. In *Ismir* (Vol. 2, pp. 287-288).
- Guichaoua, C. (2017). *Modèles de compression et critères de complexité pour la description et l'inférence de structure musicale (Doctoral dissertation, Université Rennes 1)*.
- Huron, D. (2008). *Sweet anticipation: Music and the psychology of expectation*. MIT press.
- Lerdahl, F., Jackendoff, R., Slawson, W. (1985). A Reply to Peel and Slawson's Review of "A Generative Theory of Tonal Music". *Journal of music theory*, 29(1), 145-160..
- Louboutin, C. (2019). *Modélisation multi-échelle et multi-dimensionnelle de la structure musicale par graphes polytopiques (Doctoral dissertation, Rennes 1)*.
- Louboutin, C., Bimbot, F. (2017, June). Polytopic Graph of Latent Relations: A Multiscale Structure Model for Music Segments. In *6th International Conference on Mathematics and Computation in Music (MCM 2017)* (Vol. 10527). Springer.

- Meredith, D. (2012). Music analysis and Kolmogorov complexity. In XIX Colloquio di Informatica Musicale.
- Narmour, E. (1992). The analysis and cognition of melodic complexity: The implication-realization model. University of Chicago Press.
- Narmour, E. (2000). Music expectation by cognitive rule-mapping. *Music Perception*, 17(3), 329-398.
- Neuwirth, M., Bergé, P. (Eds.). (2015). What is a Cadence?: Theoretical and analytical perspectives on cadences in the classical repertoire. Leuven University Press.
- Nika, J., Chemillier, M., Assayag, G. (2017). Improtek: introducing scenarios into human-computer music improvisation. *Computers in Entertainment (CIE)*, 14(2), 1-27.
- Peeters, G. (2007, September). Sequence Representation of Music Structure Using Higher-Order Similarity Matrix and Maximum-Likelihood Approach. In *ISMIR* (pp. 35-40).
- Raphael, C. (2002, October). Automatic Transcription of Piano Music. In *ISMIR*..
- Sadie, S., Tyrrell, J. (2001). Dictionary of music and musicians. New York: Oxford University Press. Yónatan Sánchez.
- Sargent, G., Bimbot, F., Vincent, E. (2017). Estimating the Structural Segmentation of Popular Music Pieces Under Regularity Constraints. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 344-358.
- Schenker, H. (1935). *Der Freie Satz*. Vienna: Universal Edition. Trans. E. Oster (1979) *Free Composition*, New York: Longman.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3), 379-423.
- Smith, J. B. L., Burgoyne, J. A., Fujinaga, I., De Roure, D., Downie, J. S. (2011, October). Design and creation of a large-scale database of structural annotations. In *ISMIR* (Vol. 11, pp. 555-560).
- Smith, J. B., Kawasaki, Y., Goto, M. (2019). Unmixer: An Interface for Extracting and Remixing Loops. In *ISMIR* (pp. 824-831).
- Snyder, B. (2000). *Music and memory* (p. 193). Cambridge, MA: MIT press..

Titre : Approches algorithmiques et perceptives pour la recomposition polytopique de mélodies par critères de complexité multi-échelle

Mots clés : Music Information Retrieval, Structure Musicale, Théorie de l'information

Cette thèse explore sous différents aspects musicologiques, algorithmiques et perceptifs le principe des recompositions musicales dite « polytopiques ».

Il s'agit d'un procédé de génération de variations musicales consistant à créer un nouveau contenu en réorganisant l'ordre temporel des motifs musicaux à partir d'un matériau d'origine, tout en préservant certaines propriétés de leurs relations structurelles.

Ces relations structurelles sont décrites par un graphe polytopique qui rend compte des similarités entre motifs à différentes échelles de temps, sur la base du modèle « Système & Contraste » (S&C) proposé par (Bimbot et al. 2012, 2016).

À travers une démarche créative, cette thèse vise à offrir de nouveaux paradigmes de recomposition musicale, mais également à évaluer la pertinence des hypothèses sous-jacentes au modèle polytopique, en ce qu'il rend compte de la structure musicale.

Après une présentation détaillée du modèle polytopique et de son utilisation pour recomposer des contenus musicaux, cette thèse évalue la pertinence des recompositions polytopiques par rapport à des recompositions aléatoires, grâce à un test perceptif mené sur un ensemble d'auditeurs (Gillot et al., 2019).

Une première série de tests suggère un avantage net des recompositions polytopiques vis-à-vis de recompositions aléatoires comparables en terme de perturbations de l'ordonnancement temporel.

Une seconde série de tests étudie différentes variantes d'algorithmes destinées à estimer la délimitation optimale des motifs mélodiques permettant d'obtenir des recompositions polytopiques satisfaisantes vis-à-vis de la surface musicale.

Ces algorithmes sont basés sur la minimisation d'un critère de complexité alliant le modèle Système & Contraste et des principes de Transport Optimal.

Cet ensemble de méthodes sont non seulement comparées à des annotations humaines, mais elles sont également soumises à une batterie de tests perceptifs sur un panel d'auditeurs, pour évaluer leur acceptabilité musicale.

Les résultats indiquent une relative concordance entre les niveaux d'acceptabilité des méthodes testées et des annotations produites, avec toutefois de faibles écarts entre variantes des méthodes automatiques.

Ce nouveau concept de recompositions polytopiques peut être utilisé de façon interactive à des fins créatives et un prototype pour applications mobiles a été développé en partenariat avec la PME Kercodex pour la musique au format MIDI.

Title : Algorithmic and perceptual approaches for polytopic recomposition of melodies by multiscale complexity criterion

Keywords : Music Information Retrieval, Musical Structure, Information theory

This PhD thesis explores the principles of “polytopic musical recomposition” under musicological, algorithmic and perceptual aspects.

Polytopic musical recompositions are defined as a type of musical variations which consists in creating new musical content by rearranging the temporal order of musical segments from an original material, while preserving some of its structural properties.

The structural properties are described as a polytopic graph accounting for the similarities between segments at different time scales, according to the “System & Contrast” (S&C) model proposed by (Bimbot et al. 2012, 2016).

Based on a creative approach, this work aims at proposing new paradigms for musical recomposition, but also to assess the relevance of the assumptions behind the polytopic model, as of how it accounts for musical structure.

After a detailed presentation of the polytopic model and its usability to recompose musical contents, this PhD thesis evaluates the relevance of polytopic recompositions as opposed to random recompositions, by means of a perceptual test carried out on a population of listeners (Gillot et al., 2019).

A first series of tests suggests a clear advantage of polytopic recompositions over comparable random recompositions, as regards the time-line disturbances.

A second series of tests studies a set of algorithmic variants for estimating the optimal melodic segments for polytopic recompositions, so as to optimize them with respect to the melodic surface.

The algorithms are based on the minimization of a complexity criterion relying on the S&C model combined with Optimal Transport principles.

The set of algorithmic methods are compared to human annotations, but also evaluated in the context of perceptual tests on a panel of listeners, so as to assess their musical acceptability.

Results indicate a relative agreement between the level of acceptability of the algorithmic methods versus human annotations, but they show relatively small differences between variants of automatic methods.

The new concept of polytopic recompositions can be used interactively for creative purposes, and a prototype for mobile devices has been developed for music in MIDI format in partnership with the Kercodex SME.