

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

SPÉCIALITÉ :
INFORMATIQUE

PRÉSENTÉE PAR M. BENOIT MEUDIC
POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ
PARIS 6

SUJET DE LA THÈSE :

DÉTERMINATION AUTOMATIQUE DE LA PULSATION, DE LA
MÉTRIQUE ET DES MOTIFS MUSICAUX DANS DES
INTERPRÉTATIONS À TEMPO VARIABLE D'OEUVRES
POLYPHONIQUES

SOUTENUE LE 8 Avril 2004

DEVANT LE JURY COMPOSÉ DE :
M. Gérard ASSAYAG, CO-DIRECTEUR
M. Jean-Pierre BRIOT, EXAMINATEUR
M. Jean-Marc CHOUVEL, RAPPORTEUR
M. Fabrice KORDON, EXAMINATEUR
M. Francis ROUSSEAUX, EXAMINATEUR
M. Emmanuel SAINT-JAMES, CO-DIRECTEUR
M. Gerhard WIDMER, RAPPORTEUR

Remerciements

Pour être sur de n'oublier personne, je remercie par avance tous ceux que je vais oublier, une fois passée cette phrase, et grâce à qui la vie me paraît chaque jour plus belle (j'hésite à remercier les autres).

Plus particulièrement, je voudrais remercier Pierre Lusson et François Sarhan pour leur initiation à la fameuse 'théorie du rythme de Pierre Lusson', qui même si elle n'a pu se concrétiser par un travail de thèse faute de moyens, m'a fortement inspiré au long des trois années de la présente.

En fait, c'est à l'union européenne que je dois mon financement, et je tiens particulièrement à remercier Hugues Vinet pour le contrat et les conditions de travail qu'il m'a offertes au sein du projet Cuidado.

Pour l'ambiance détendue et amicale de l'équipe rep-mus, qui m'a accueilli pendant ces trois années, je remercie Gérard, Moreno, Carlos, Karim, Jean et Charlotte, ainsi que tous les stagiaires pour les parties de Tennis, d'Echecs ou de Natation.

La natation me fait d'ailleurs penser à Francis, amateur de baies nautiques dont celles de Naples, Marseille et Les Halles.

Ce travail de thèse n'aurait pu se faire sans le soutien et les conseils avisés d'Olivier Lartillot, qui tout au long de ces trois années m'a accompagné à la fois pendant les pauses café et les périodes de réflexion intenses, lui-même réalisant de son côté une thèse dont le sujet n'est pas sans rapports avec le mien.

A sujets différents mêmes directeurs, je remercie donc Gérard Assayag et Emmanuel Saint-James pour leur co-direction de cette thèse, toujours atten-

tifs à m'accompagner dans les moments délicats et prompts à la discussion lorsque le besoin s'en fait sentir. Par la même occasion, je remercie aussi les autres membres de mon jury pour m'avoir fait l'honneur d'accepter cette fonction, messieurs Briot, Kordon et Rousseaux, et particulièrement mes deux rapporteurs Jean-Marc Chouvel et Gerhard Widmer.

Pour leur relecture entière du tapuscrit, et leur expertise de bricoleurs avertis grâce à laquelle j'ai pu consacrer un peu plus de temps à cette thèse, je remercie les parents d'Odile, Elisabeth et Jean.

Mes parents Isabelle et Alain pour m'avoir toujours soutenu, alors même que je m'éloignait du terreau familial de la médecine pour m'aventurer sur le terrain vague de la recherche. Ma sœur d'ailes fines et mon frère Yann ainsi que sa petite famille.

Et Odile.

Résumé

Les questions que pose l'analyse musicale assistée par ordinateur sont souvent abordées tantôt sous leur aspect technique, lorsque l'analyste se dit informaticien, tantôt sous leur aspect musical, lorsque l'analyste se dit musicologue. A un même problème correspondront alors des réponses différentes. Nous pensons qu'une utilisation de l'informatique en analyse musicale doit tirer parti de ces deux aspects plutôt que de les opposer. Une analyse ne fera alors appel à la rationalité informatique ou à un discours musicologique que si cette rationalité ou ce discours sont réellement requis.

Nous proposons dans cette thèse un des seuls programmes à notre connaissance d'analyse rythmique et motivique automatique de musiques polyphoniques. Nous répondons au problème central de la complexité combinatoire par une exploration progressive du texte musical à partir de son niveau symbolique le plus bas. Nous partons de la note pour extraire une pulsation et une métrique sur lesquelles nous prenons ensuite appui pour analyser les variations motiviques. Notre méthode s'applique aussi bien aux monodies qu'à la musique polyphonique, ne se restreint pas aux métriques binaire ou ternaire, et détecte les variations de tempo. Elle ne repose sur aucun système harmonique présumé (tonalité, modalité ou autre), n'étant limitée que par la représentation symbolique du tempérament égal employé par le format MIDI. Le corpus visé est donc large, nos exemples partant du classicisme jusqu'au jazz post-bop en passant par l'atonalité expressionniste. Le critère de validité de nos résultats est alors leur cohérence perceptive.

Les avantages d'une telle approche sont nombreux. D'une part, nous

sommes amenés à nous poser des questions sous un angle inhabituel. L'analyse étant réalisée par une machine, il nous faut externaliser non plus simplement les éléments que l'on perçoit, notés par exemple sous forme de partition, mais aussi les mécanismes qui nous amènent à leur perception. Cette mise en perspective sera alors décisive pour mieux cerner certaines problématiques liées au concept de polyphonie ou de pulsation. Par ailleurs, le fait d'analyser des musiques complexes donc réalistes permet d'envisager plusieurs applications immédiates, comme la recherche de motifs saillants au sein d'une polyphonie, la recherche des variations d'un motif donné à l'intérieur d'une grande base de données musicales, ou la comparaison entre différentes interprétations rythmiques d'un même morceau. Enfin, les premiers résultats montrent que certaines structures musicales jusque-là perçues de façon implicite peuvent être exhibées par notre système, ce qui nous laisse augurer la possibilité d'une nouvelle forme d'intelligence avec l'expressivité musicale.

Abstract

The issue of computer-assisted musical analysis is often tackled either under its technical aspect, when the analyst is said to be a computer scientist, or under its musical aspect, when the analyst is said to be a musicologist. To the same problem will then correspond different answers. We think that the use of computer science in musical analysis must rather take advantage of the two aspects than oppose them. An analysis will then require computer rationality or musicologic speech only if this rationality or this speech is really necessary.

We propose in this phd one of the few programs, to our knowledge, that can automatically analyse rhythm and motives from polyphonic music. Our answer to the central problem of the combinative complexity is to explore progressively a musical piece from its lower symbolic level. Starting from the note, we extract the beat and the meter on which we rely for analysing motivic variations. Our method is convenient for monodies as well as polyphonies, is not restricted to binary or ternary meter, and is able to detect tempo variations. It does not rely on any presupposed harmonic system, and is only limited to the symbolic representation in MIDI of the equal temperament. The range of music that can be analysed start from classical to jazz post-bop and expressionist atonality. The criteria for the validity of the results is their perceptive relevance.

Such an approach offers several advantages. First, questions are raised from a new point of view. The analysis being performed by the machine, we have to externalize not only what we perceive, that can be represented

for instance by scores, but also the mechanisms of that perception. This new perspective will help us to better delimit fundamental concepts related to polyphony or beat. Moreover, the consideration of complex music makes our algorithms, such as pattern extraction, search for pattern variations or performance comparison, directly suitable for real pieces of music. Last, the first results show that some musical structures that used to be perceived but not explicated can be extracted by our system, which let us believe that a new form of intelligence of musical expressivity is possible.

Table des matières

1	Introduction	1
1.1	Analyse et musique	1
1.2	L'outil informatique dans l'analyse	3
1.3	Objectifs	4
1.4	Organisation	4
2	Les sciences au service de l'analyse musicale	7
2.1	Une multiplicité d'analyses possibles	8
2.1.1	L'analyse musicale, une science ambiguë ?	8
2.1.2	Le quoi, le comment et le pourquoi	9
2.2	Les modèles théoriques en analyse musicale pré-computationnelle	11
2.2.1	L'influence des mathématiques	11
2.2.2	L'influence des compositeurs	12
2.2.3	Les théories issues de la linguistique	13
2.2.4	L'analyse paradigmatique	17
2.2.5	La perception, un critère récurrent en analyse	20
2.3	Les modèles théoriques en analyse computationnelle	23
2.3.1	Temperley	23
2.3.2	Conklin	24
2.3.3	La GCTMS de Cambouropoulos	25
2.4	Synthèse	26
2.4.1	Quelle méthodologie en analyse musicale ?	26
2.4.2	Deux opérations de base pour la recherche de structures	31

2.4.3	L'informatique est-elle nécessaire à l'analyse musicale ?	34
2.5	Notre modèle d'analyse	38
2.5.1	Le matériau de départ	39
2.5.2	Le rythme, un composant de base	39
2.5.3	La recherche des répétitions motiviques	40
3	L'analyse rythmique : le cas de la pulsation	41
3.1	Qu'est-ce qu'un rythme ?	41
3.1.1	Un invariant : la temporalité	42
3.1.2	Le rythme un phénomène immatériel ?	44
3.1.3	Comment perçoit-on un rythme dans le temps ?	45
3.1.4	Nos représentations du rythme	46
3.1.5	Conclusion	47
3.2	Extraire la pulsation	47
3.2.1	Définitions	47
3.2.2	Les différentes approches	55
3.2.3	Un modèle d'accentuation des attentes	57
3.2.4	Notre algorithme d'extraction	62
3.3	Conclusion : principaux avantages par rapport à la méthode de Dixon	73
4	L'analyse rythmique : métrique et quantification	77
4.1	La mesure	77
4.1.1	Définitions	77
4.1.2	Les différentes approches	81
4.1.3	Conjuguer l'analyse des similarités avec celle des accents	82
4.1.4	Première étape, rechercher une structure d'accentuation	83
4.1.5	Deuxième étape, rechercher un groupement	85
4.1.6	La recherche d'un groupement : analyser les répétitions par autocorrélation	86
4.2	La quantification des durées en intervalles proportionnels . . .	91
4.2.1	Définitions	91

4.2.2	Quelques modèles de quantification	92
4.3	Exemples	93
4.4	Récapitulatif	97
5	L'analyse motivique : la mesure des similarités	99
5.1	Qu'est-ce qu'un motif ?	100
5.1.1	La répétition comme invariant	100
5.1.2	Le motif, une notion perceptive	101
5.1.3	Comment modéliser un motif polyphonique ?	103
5.2	Différentes algorithmes d'analyse motivique	107
5.2.1	L'analyse de fichiers audio	107
5.2.2	La complexité informative	107
5.2.3	La programmation dynamique	108
5.2.4	Une approche issue de la linguistique	109
5.2.5	La modélisation des accents	109
5.2.6	La visualisation dans l'espace	110
5.2.7	La modélisation du processus d'induction	111
5.2.8	La similarité par partage de propriétés communes . . .	111
5.2.9	Synthèse	113
5.3	Extraction des épures de deux séquences polyphoniques	116
5.3.1	Une notion de départ : l'intersection	117
5.3.2	Les règles cognitives	121
5.3.3	La similarité par contour de hauteurs	124
5.3.4	La similarité par intervalles de hauteur horizontaux . .	129
5.3.5	La similarité par intervalles de hauteur verticaux . . .	134
5.3.6	La similarité rythmique	138
5.3.7	La similarité globale	141
5.3.8	La représentation en épures, ou la synthèse par l'analyse	143
5.4	Récapitulatif	143
6	L'analyse motivique : l'extraction et la recherche de motifs	147
6.1	Représentation matricielle des similarités	148

6.1.1	Pourquoi fixer la taille des séquences à comparer ? . . .	150
6.1.2	Choisir le bon niveau	152
6.1.3	Le calcul d'une matrice - synthèse	156
6.2	L'extraction de motifs à partir de la matrice de similarité . . .	160
6.2.1	Quels sont les motifs 'les plus importants'	160
6.2.2	Commencer par filtrer les cellules de la matrice	161
6.2.3	Etablir une classification des cellules d'une même ligne	162
6.2.4	Comment extraire les motifs du plus petit au plus grand	165
6.2.5	Comment extraire les motifs tout court	167
6.2.6	Définir les débuts et fins de motifs	168
6.3	L'extraction de motifs à partir d'une séquence de référence . .	169
6.3.1	Une complexité moindre	169
6.3.2	Préserver une cohérence rythmique	171
6.3.3	Un outil pour établir des liens entre différents morceaux	172
6.4	Récapitulatif	174
7	Synthèse et exemples	177
7.1	Le modèle général d'analyse	177
7.1.1	Formatage du fichier MIDI	178
7.1.2	Recherche d'une pulsation	179
7.1.3	Extraction de la métrique	181
7.1.4	Mesure des similarités	182
7.1.5	Calcul de la matrice de similarité	182
7.1.6	Extraction de motifs représentatifs et recherche de liens	183
7.2	Analyse du début de la 'Sonate n°9' de Beethoven	183
7.2.1	Recherche d'une pulsation	183
7.2.2	Recherche d'une mesure	184
7.2.3	Calcul de la matrice de similarité	186
7.2.4	Extraction motivique	188
7.3	Le début de la 'Pavane pour une infante défunte' de Ravel . .	189
7.3.1	Recherche d'une pulsation	191
7.3.2	Calcul de la matrice de similarité	191

7.3.3	Extraction motivique	193
8	Conclusion	197
8.1	Deux librairies informatiques	197
8.2	Amélioration de l'existant	198
8.3	Travaux futurs	199
ANNEXES		201
A	Documentation du logiciel MusicMap	203
A.1	Introduction	203
A.2	General description of the library	204
A.3	Pattern extraction	205
A.3.1	Global description	205
A.3.2	Automatic search of repetition	211
A.3.3	Extracting patterns from the matrix	214
A.4	Search for a polyphonic pattern in a MIDI database	215
A.4.1	Global description	215
A.4.2	More details on the similarity measure	217

Chapitre 1

Introduction

« Si un contemplatif se jette à l'eau, il n'essaiera pas de nager. Il essaiera d'abord de comprendre l'eau. Et il se noiera. » (Henri Michaux, Extrait de *Le portrait de A, Difficultés, Lointain intérieur*)

1.1 Analyse et musique

Devant la quantité d'informations qui nous parviennent quotidiennement, l'analyse devrait être une activité centrale dans notre compréhension des phénomènes de ce monde. A quoi sert-il de lire, écouter ou regarder les médias si l'on ne procède pas systématiquement à une analyse des informations que l'on choisit de nous livrer ? Comment comprendre sinon un tel flux d'informations ? Outil de pouvoir, mais aussi d'émancipation individuelle, l'analyse est une activité souvent présente mais cachée à différents niveaux de notre société. Stratégies de communications, discours politiques, politiques économiques sont le fruit d'analyses mûrement réfléchies mais presque invisibles pour ceux qui n'en sont pas les auteurs. Alors que dans un passé lointain l'instinct était notre principal instrument de défense face aux dangers du monde extérieur ou intérieur, nous prévenant par exemple de la noyade ou de la déshydratation, l'activité d'analyse s'avère de plus en plus nécessaire pour se prémunir des nouvelles offensives transmises par le vecteur de la

(dés)information.

Qu'en est-il du domaine artistique, et plus particulièrement de la musique ? L'activité d'analyse ne concentre pas les mêmes enjeux, et s'envisage de façon plus sereine, mais elle nous offre toujours un moyen de comprendre, à travers la musique, les phantasmes, obsessions et croyances qui ont guidé et guident toujours notre destinée d'êtres humains. Pour s'en tenir à la musique, l'analyse nous fournit certaines clefs qui nous permettent d'appréhender l'essence même du phénomène musical. Son rôle dans l'activité d'écoute peut de prime abord paraître ambiguë : est-il nécessaire par exemple d'analyser la musique contemporaine pour pouvoir l'écouter convenablement, ou au contraire nous est-elle immédiatement accessible, sans l'intermédiaire de l'analyse ? Peut-être est-ce une fausse question, puisque comme le dit Jean-Michel Bardez dans sa préface de *l'analyse musicale* [3] :

« Ecouter une œuvre, c'est 'l'interpréter', l'analyser, de quelque manière que ce soit. Plus dense, plus diversifiée sera notre ouverture à l'entendement de l'œuvre, plus ramifiés, plus subtils seront nos moyens de l'appréhender, plus intenses seront nos plaisirs, émotions et compréhensions, 'saveurs' et 'savoir' toujours liés en une série d'actes relatifs très complexes. »

Ainsi, l'analyse est présente dès qu'il s'agit d'appréhender l'œuvre musicale. Par ailleurs, elle diffère à chaque écoute, pour chaque 'entendeur' :

« C'est l'infinie pluralité des écoutes qui constitue l'œuvre [...] C'est 'son' écoute [à soi], celle d'un autre 'entendeur', signifiée, relatée, exprimée en mots, en images, en systèmes, en parcours, en tracés, en 'découpes', en diagrammes, en chiffres, en dessins, en passions même. »

L'analyse est multiple, c'est l'expression de la réception de l'œuvre par chacun suivant son langage, ses émotions, ses attentes. Par l'analyse, chacun communique sa propre vision de l'œuvre. Comment alors, face à cette diversité, imaginer que certaines théories puissent émerger, s'imposer comme des modèles acceptés par plusieurs 'entendeurs'. Y a-t-il des écoutes plus

vraies que les autres, plus à même d'intégrer d'autres écoutes, ou faut-il plutôt supposer l'existence de règles universelles que l'analyse aurait pour tâche de découvrir ? Comment imaginer un outil informatique capable lui aussi de proposer une écoute qui fasse sens pour nous ?

Ce sera l'objet de cette thèse, que nous placerons donc au point de rencontre entre les sciences informatiques et l'analyse musicale.

Avant de préciser la méthodologie et l'organisation choisies pour cette thèse, nous allons entrer plus en avant dans le domaine informatique.

1.2 L'outil informatique dans l'analyse

Auteur d'une thèse récente portant sur la construction autonome des représentations, Rispoli [42] introduit son sujet par le constat que :

« L'intelligence artificielle est devenue ces dernières années une science pour l'ingénieur à part entière. En particulier, il a été considéré comme acquis qu'il fallait se concentrer sur les difficultés algorithmiques plutôt que sur l'origine des objets manipulés. »

Rispoli remet alors en cause ce constat en argumentant que :

« Un prétraitement qui consiste à identifier les objets d'étude par le système lui-même est absolument primordial et non pas secondaire car il détermine la forme de ces objets et ce qui en sera déduit ultérieurement. »

Replacé dans le contexte de l'analyse musicale, ce constat reste toujours aussi vrai. Nous pensons que les utilisations actuelles de l'informatique en analyse musicale ne prennent pas suffisamment en compte l'aspect spécifiquement musical de l'objet analysé, ce qui se traduit par une focalisation de la recherche sur le traitement des données, faisant intervenir des algorithmes complexes, sophistiqués, mais inadaptés au problème musical, au détriment d'une étude approfondie des schémas de représentation à partir de laquelle seulement nous pourrions espérer obtenir certains résultats.

1.3 Objectifs

Nous pouvons résumer nos objectifs en trois points clef :

- L’objet d’étude : on choisit d’analyser deux composants très importants de la musique que sont le rythme et les motifs ;
- Le matériau : on choisit d’analyser ces composants pour des musiques complexes (musique polyphonique, de différentes époques, de styles différents) en portant notre attention sur la nature des objets étudiés ;
- L’outil utilisé : on choisit de réaliser un outil informatique le plus général possible nécessitant un minimum d’interventions de la part de l’utilisateur.

Au vu des objectifs, nous nous trouvons dans une situation inédite pour laquelle aucun des logiciels actuels d’analyse musicale ne propose de réponse. Notre système sera donc le premier à notre connaissance, à proposer l’extraction automatique de motifs polyphoniques à partir de musiques symboliques polyphoniques. Notons que si notre logiciel peut traiter la musique polyphonique, il sera aussi naturellement applicable à la musique monophonique.

1.4 Organisation

Le document se divise en huit chapitres y compris celui-ci et la conclusion.

Le chapitre deux est consacré à la présentation des théories et techniques par rapport auxquelles s’inscrit notre travail, ainsi qu’à la présentation générale de notre logiciel d’analyse.

Le chapitre trois aborde la question de l’analyse automatique du rythme et plus particulièrement celle de la pulsation à partir d’une modélisation de la notion musicale d’attente et de la détection des accents rythmiques sous l’angle de la théorie de la gestalt. Le chapitre quatre est axé sur l’analyse du mètre par autocorrélation et sur la quantification, ces deux traitements s’effectuant à partir de la donnée d’une pulsation.

Le chapitre cinq prend aussi comme hypothèse de départ la donnée d'une pulsation pour proposer une analyse automatique de la similarité entre deux séquences MIDI basée sur la considération de trois points de vue que nous jugeons fondamentaux du phénomène musical.

Le chapitre six prend comme hypothèse la connaissance d'une mesure de similarité pour tenter de résoudre deux problèmes concrets : le premier concerne l'extraction automatique de motifs à partir d'une matrice de similarité calculée sur un fichier MIDI, le deuxième concerne la recherche des variations d'un motif à l'intérieur d'une base de données musicale.

Enfin, le chapitre sept synthétise l'ensemble des résultats des chapitres précédents à travers différents exemples musicaux et le chapitre huit propose une conclusion.

Chapitre 2

Les sciences au service de l'analyse musicale

L'analyse musicale n'est pas une discipline scientifique. Pourtant, de plus en plus, et au fur et à mesure des avancées technologiques, elle utilise les sciences dans une tentative toujours plus exigeante de formaliser le phénomène musical. Cette tendance à l'exigence est accentuée par le nouveau potentiel d'objectivation qu'offre l'informatique, qui permet d'expérimenter concrètement par le calcul les différents modèles proposés par les théories. Par ailleurs, l'informatique offre à la réflexion musicale de nouveaux concepts dont le potentiel n'est pas encore clairement délimité et qui pourraient être à l'origine d'un renouveau de la pensée analytique.

A partir d'une présentation des modèles théoriques les plus influents en analyse musicale, pour laquelle nous marquerons une distinction entre modèles computationnels et modèles non computationnels, nous détaillerons dans ce chapitre les motivations à l'origine de notre nouveau programme d'analyse, en portant l'accent sur le rôle particulier que joue l'informatique. Mais tout d'abord, nous allons délimiter plus précisément le domaine de l'analyse musicale que nous nous proposons d'explorer.

2.1 Une multiplicité d’analyses possibles

La première difficulté de tout travail d’analyse vient de la nécessité de définir précisément des objectifs au regard de la multiplicité des chemins possibles d’exploration. La reconstitution des étapes de la composition d’une œuvre, la recherche de singularités dans le matériau musical ou la mise en perspective historique de différentes interprétations d’une même œuvre, sont autant de pistes d’exploration possibles qui relèvent du domaine de l’analyse musicale sans pour autant que les méthodologies requises ne soient communes. Au regard des objectifs que nous nous sommes fixés au premier chapitre, nous allons donc définir le champ de l’analyse musicale qui nous intéresse, en commençant par adopter dans la section suivante une définition proposée par Ian Bent qui place l’analyse dans un idéal cartésien, qui, même s’il ne peut en tant qu’idéal être atteint, fournit l’arrière-plan des motivations qui nous guideront dans l’élaboration de notre modèle.

2.1.1 L’analyse musicale, une science ambiguë ?

Au début de son ouvrage *l’analyse musicale* [3], Ian Bent propose la définition suivante, que nous adopterons par la suite :

« L’analyse musicale est la ‘résolution’ d’une structure musicale en éléments constitutifs relativement plus simples, et la recherche des ‘fonctions’ de ces éléments à l’intérieur de cette structure. »

Cette définition montre bien l’ambiguïté inhérente à l’analyse musicale, qui vise idéalement à élucider par des termes scientifiques un problème qui néanmoins dépasse le simple fait scientifique : faut-il ‘résoudre’ une structure musicale comme on résoudrait une équation ? Faut-il formuler les ‘fonctions’ des notes et accords à l’intérieur d’une séquence musicale comme on formulerait des fonctions à l’intérieur d’un espace géométrique ?

La réponse est immédiatement négative, puisqu’il n’existe pas une seule analyse possible d’une œuvre (une seule solution), mais plusieurs, et modé-

liser les différents problèmes analytiques par autant d'équations à résoudre serait peut-être une tâche plus difficile que le fait d'y répondre directement.

Cependant, une telle ambiguïté présente des avantages : voulue comme une science, mais lui échappant, l'analyse musicale reste un domaine à conquérir, et donc stimule la recherche scientifique.

Inversement, devant l'infinité des interprétations potentielles d'une même œuvre musicale à laquelle pourrait conduire une analyse purement esthétique et empirique, la formalisation scientifique a la particularité de proposer une tentative d'unification et de généralisation de nos expériences individuelles d'écoute. Cet aspect pourrait être critiqué par les tenants d'une analyse singulière propre à chaque analyste mais cette critique serait fondée sur un malentendu. En effet, s'il ne faut pas tomber dans le piège de l'uniformisation qu'entraînerait la recherche d'un unique point de vue, nous pensons qu'une analyse totalement empiriste ne pourrait inversement être partagée par plusieurs, et n'aurait d'intérêt en tant qu'analyse que pour celui qui la réalise. Une analyse peut être singulière ou non, c'est à dire propre ou non à la personne qui l'a effectuée, mais elle n'aura de sens pour nous que si les règles et concepts ayant conduit à sa réalisation nous sont soit communs, soit explicités, afin que nous puissions nous les approprier. Dans le cas contraire, l'analyse sera considérée dans le meilleur des cas comme une œuvre d'art, comme un objet dont l'appropriation pourra alors se faire par l'intermédiaire d'une nouvelle analyse (une méta-analyse).

2.1.2 Le quoi, le comment et le pourquoi

A partir de la formulation en trois questions de l'analyse musicale par Jean Marc Chouvel [2] nous pouvons tenter de circonscrire davantage le domaine qui nous intéresse.

Jean Marc Chouvel pose le problème de l'analyse musicale en trois questions :

- le quoi, ou la question du matériau ;
- le comment, ou la question de l'organisation du matériau ;

- le pourquoi, ou la question des raisons externes au quoi et au comment.

La question du matériau concerne l'objet sur lequel porte l'analyse. Sans entrer dans les détails, qui seront explicités par la suite, nous avons déjà choisi de nous intéresser uniquement à l'œuvre sous forme de partition ou sous d'autres formes de représentation similaires, par exemple au format MIDI. MIDI signifie Musical Instrument Digital Interface. Ce format permet entre autre de coder la musique sous forme de séquences de hauteurs, durées, onsets et intensités (le terme onset désigne le temps d'attaque d'une note. Il est souvent exprimé en millisecondes). On peut donc exclure de notre analyse les autres traces de la composition laissées par le compositeur telles que les feuillets, notes, dessins, algorithmes ou patches, utiles par exemple en analyse génétique. Les textes, critiques ou éléments bibliographiques du compositeur seront aussi laissés de côté tout comme les analyses ou commentaires effectués par un tiers.

Le comment, si l'on adopte la définition de l'analyse musicale par Ian Bent, pose la question de la structure. Qu'est-ce qu'une structure, comment est-elle organisée? Si de manière générale chercher la structure d'un objet revient à expliciter l'organisation de ses éléments constitutifs, la musique pose le problème spécifique du temps. Rechercher une structure musical revient donc à étudier comment les différents éléments d'une séquence sont organisés dans le temps. Nous aborderons ce problème dans la section suivante à travers la présentation des principaux courants d'analyse, avant de proposer notre propre approche.

Le pourquoi concerne les raisons externes du quoi et du comment, par exemple "quelles sont les motivations du compositeur ayant conduit à l'œuvre musicale"? C'est aussi la question du sens, souvent laissée de côté car difficile à formaliser. Par contre, les motivations du compositeur nous semblent importantes pour décider de l'angle d'analyse. Par exemple, on peut penser que tout compositeur, destinant sa musique à l'écoute, utilisera plus ou moins consciemment des critères perceptifs dans l'élaboration de son œuvre. Ces critères pourraient donc nous guider dans l'étude de l'organisation du mor-

ceau. Nous essaierons d'utiliser cet aspect en argument de notre algorithme d'analyse, cette question ne faisant pas l'unanimité parmi toutes les théories analytiques que nous allons présenter.

2.2 Les modèles théoriques en analyse musicale pré-computationnelle

2.2.1 L'influence des mathématiques

Parmi les domaines ayant contribué au développement de la pensée musicale, les mathématiques figurent parmi les plus influents. Toutefois, leur contribution au domaine de l'analyse musicale reste difficile à définir précisément.

L'une des problématiques concerne le rapport possible entre logique musicale et logique mathématique. Ainsi, François Nicolas en introduction du dernier forum Diderot Mathématique-Musique [14] se demande :

« [...] les formalismes bâtis sur le 'raisonnement' musical (l'ensemble emmêlé de ses rationalités, décisions, buts, et de ses calculs, déductions, enchaînements) ont-ils quelque chose à voir avec tel ou tel formalisme de la logique ? Quelle cohérence formelle, logiquement descriptible, peut-elle exister dans l'œuvre, par-delà l'arbitraire des décisions esthétiques singulières ?[...] »

Une réponse immédiate est que les formalismes de la logique, et des mathématiques en général, ne peuvent s'appliquer directement à la musique puisque les critères décidant du vrai ou du faux diffèrent entre ces deux domaines. Par exemple, Mikhail Malt [27] remarque que sur l'application en musique du concept mathématique de 'relation' :

« [...] musique et mathématiques, chacune valide et traite les relations de façons différentes. Pendant que les mathématiques valident la cohérence des relations à partir de démonstrations formelles, la musique valide ses relations à partir d'une conjonction,

très floue, de critères rationnels, formels, esthétiques et sensibles.

»

Cependant, certains aspects musicaux bien choisis peuvent être étudiés en termes mathématiques, et inversement, la musique peut rendre compte de certains phénomènes mathématiques.

Si l'on s'en tient au domaine de l'analyse musicale, la 'set-theory' est souvent présentée comme étant la théorie 'mathématique' de la musique la plus complètement formalisée. La 'set-theory' repose sur la notion mathématique d'ensemble que l'on applique aux notes de la partition et sur lesquelles on peut appliquer certaines des opérations de la théorie des ensembles choisies en fonction de leur sens musical. Par exemple, la transposition, la symétrie par rapport à la note la plus grave ou le renversement sont des opérations très utilisées. Dans une première étape, les notes de la partition doivent être regroupées en ensembles selon un processus non formulé par la théorie, laissé au libre choix de l'utilisateur. Ainsi, une multitude d'analyses toutes aussi valables les unes que les autres sont possibles.

Un autre aspect mathématique parfois utilisé en analyse musicale relève des statistiques et probabilités. Par exemple, Cope [10] propose un modèle d'analyse du style basé sur la classification statistique des segments les plus souvent répétés, et l'élaboration d'un système de règles probabilistes d'enchaînement de notes établies à partir d'une base d'exemples. L'idée générale est que la redondance, donc la confirmation de l'attente, ou la non redondance, donc la non information, jouent un rôle particulier en créant de la forme et de la structure.

2.2.2 L'influence des compositeurs

Si l'on peut difficilement contester l'existence d'une frontière entre analyse et composition, il s'agit d'une frontière perméable laissant filtrer les idées et concepts d'un domaine à l'autre. Devant la multiplicité des principes pouvant être mis en œuvre dans une composition, l'analyse s'efforce de trouver un langage plus général, donc moins riche, pouvant à lui seul rendre compte de

plusieurs compositions. Si ce langage ne peut être celui d'un compositeur en particulier, il reste toujours très instructif de considérer les remarques sur l'analyse musicale faites par ceux qui sont à l'origine des pièces étudiées.

Plusieurs études s'intéressent au rapport entre l'analyste, le compositeur et l'œuvre. Par exemple, Fabien Levy [23] note que les transformations perçues dans une œuvre peuvent être très éloignées de celles mises en œuvre dans le processus compositionnel ce qui argumente pour une définition précise de l'objet sur lequel porte l'analyse, souvent laissé dans le flou par les théories.

Inversement, des théories compositionnelles peuvent être mises en lumière par l'analyse d'une œuvre elle-même initialement composée par d'autres principes que ceux utilisés dans l'analyse. On pourrait par exemple effectuer une analyse sérielle, stochastique ou algébrique d'une sonate de Beethoven.

Il ressort alors qu'analyse et composition sont étroitement liés, l'un pouvant mettre en lumière les idées de l'autre dans un processus d'échange et d'enrichissement mutuel. Ainsi, le texte constituant l'analyse nous éclaire parfois autant sur l'œuvre analysée que sur l'analyste. Inversement, l'œuvre peut parfois tirer son matériau d'une réflexion sur elle-même, comme l'illustre cette réflexion d'Helmut Lachenmann [19] :

« [Suite à la prise de position structuraliste des années 50,]
la musique perdit son caractère de langage et se découvrit n'être
qu'une structure muette - mais non moins éloquente et toujours
chargée d'expression, bien que de manière moins agréable. [...] Dans
cette phase historique, pour 's'exprimer', la musique ne visait pas à rendre sa structure consciente, mais, plus radicalement, elle faisait du renouvellement du concept de matériau la matière même de l'invention de l'œuvre. La musique pensait le phénomène en lui-même [...] »

2.2.3 Les théories issues de la linguistique

La question du langage en musique a souvent été associée aux problématiques posées par la linguistique. Si le débat ne porte plus sur le primat

originel d'un domaine sur l'autre, qui a été par exemple discuté par Rousseau [45], la terminologie linguistique, souvent employée lorsqu'il s'agit d'analyser la musique, ou inversement les nombreux termes musicaux communs à la poésie, tels que refrain, couplet, mètre ou accent montrent bien que les deux domaines présentent de nombreuses similarités.

On est alors tenté de faire un rapprochement entre les questions posées dans un domaine et les réponses apportées par l'autre.

Les différentes théories d'analyse musicale inspirées de la linguistique qui ont émergé au cours du siècle dernier se rejoignent dans une tentative d'exprimer la structure d'une séquence musicale de la même manière que pourrait l'être celle d'un texte, c'est à dire en proposant à partir d'un vocabulaire de base constitué de quelques éléments une abstraction de la séquence sous forme de schéma syntaxique et/ou grammatical.

Le schéma d'abstraction de la séquence est souvent déterminé a priori, ce qui peut apparaître comme une contrainte assez forte sur le type de séquence pouvant être analysé. Pourtant, de nombreuses théories (par exemple les théories fort-faible qui seront présentées dans la section suivante) utilisent ce principe et ont eu une grande influence sur l'analyse musicale.

Certaines théories, comme les grammaires génératives, proposent un modèle plus souple, basé sur des règles locales d'enchaînement et un vocabulaire de base, sans définir a priori de schéma global. Les règles étant appliquées localement, un plus grand nombre de formes musicales peut a priori être analysé.

Les théories fort-faible

A la fin du XIXème siècle, Riemann [41] émet comme hypothèse que toute structure rythmique est schématisable par la composition à différents niveaux de la séquence iambique faible-fort appelée 'Takt'. Le mot rythme ne doit pas être réduit aux seules notions d'onsets et de durées, mais désigne plus généralement l'organisation de tous les paramètres musicaux dans le temps. Au niveau le plus bas, les deux éléments faible-fort du 'Takt' sont

appelés upbeat et downbeat, puis successivement mesure faible-forte, groupe de mesures faible-fort, moitié de phrase faible-forte, phrase faible-forte et période.

Pour Riemann, une structure se divise donc toujours en deux moitiés, l'une faible l'autre forte et ainsi de suite jusqu'à atteindre le niveau upbeat-downbeat, dont chacun des deux composants est lui-même un groupement de une ou plusieurs notes. Les structures rythmiques qui ne répondraient pas à ce schéma en seraient dérivées par le biais de transformations que Riemann a par ailleurs formalisé, telles que l'élision, la répétition cadentielle, le recouvrement partiel de deux phrases consécutives, l'anacrouse générale ou le motif annexé. Lidov [24] note qu'il serait alors intéressant d'étudier si musicalement de telles figures transformées seraient en effet perçues comme plus complexes que des figures répondant au schéma de base.

De même que Riemann, Cooper et Meyer [11] proposent de décrire une structure rythmique (au sens général du mot rythme) à partir de la notion d'accent fort - accent faible. Plutôt que de ne considérer que le groupement iambique, ils définissent cinq groupements abstraits différents, par ailleurs fréquemment utilisés en linguistique :

- iambe (- *);
- anapeste (- - *);
- trochée (* -);
- dactyle (* - -);
- amphibraque (- * -).

Le résultat est une segmentation de la séquence musicale en différentes couches hiérarchisées de groupements rythmiques abstraits. Les types de groupements et la détermination des accents se fait de manière intuitive au cas par cas, Cooper et Meyer se laissant guider par leurs propres expériences perceptives.

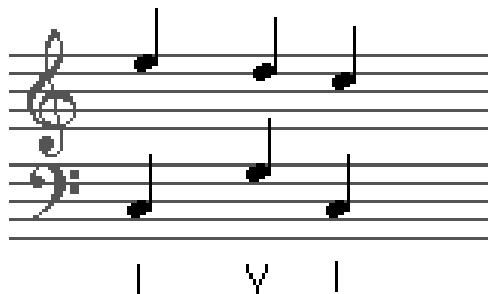


FIG. 2.1: La structure fondamentale, ou ‘ursatz’ selon Schenker.

Les théories de l’élément unique

Au début du XXème siècle, Schenker [48], au contraire de Riemann ou Cooper et Meyer, pose le problème des frontières entre les groupes, et délaisse le discours verbal pour le graphique. Par ailleurs, la dimension rythmique relative aux durées et onsets (positions d’attaque dans le temps des événements musicaux) est délaissée au profit de la mélodie et de l’harmonie. Pour Schenker, l’analyse consiste à déterminer une structure fondamentale, qu’il appelle ‘ursatz’ (voir la figure 2.1), autour de laquelle se déploie l’œuvre musicale.

Plutôt qu’une juxtaposition de groupements, l’œuvre est donc vue comme une seule entité dynamique, sorte d’épine dorsale reposant sur quelques notes et fonctions harmoniques essentielles autour desquelles viennent se greffer des éléments d’ornementation et de variation apportant à l’œuvre sa singularité. Pour la représentation des résultats, Schenker utilise une partition en superposant les différentes couches de l’analyse sur autant de portées. De même que pour les théories précédentes, le processus de réduction n’est pas formalisé, et repose en grande partie sur les intuitions de l’analyste.

A l’inverse, Réti [40] considère que l’œuvre musicale repose sur quelques motifs de base qui, après avoir été exposés, sont variés et transformés suivant des opérations qu’il se propose de formaliser. Cette vision se rapproche de la notion d’improvisation.

La ‘Generative theory of tonal music’

Une théorie générative utilise une fonction ou un ensemble de règles pour construire des objets reproduisant l’ensemble des objets à analyser. Si cette théorie s’avère pouvoir générer plus d’objets qu’il n’y en a contenus dans l’ensemble à analyser, on dit que la théorie est sur-générative. Dans le cas inverse, c’est à dire si elle ne peut générer que quelques objets parmi tous ceux à analyser, elle est sous-générative.

En musique, la *Generative theory of tonal music* de Lerdahl et Jackendoff [22] pourrait alors plutôt être appelée ‘théorie sur-générative’. En fait, il serait même théoriquement possible de s’en servir pour générer toutes les pièces de musique possibles, puisque ses règles ne forment pas contrainte sur la musique à analyser. La théorie en elle-même a eu beaucoup d’influence en analyse musicale, notamment parce qu’elle tente de formaliser pour la première fois les intuitions d’un auditeur musicien à l’aide des outils assez en vogue à l’époque de la linguistique. A partir d’une partition, différentes structures sont proposées : ‘grouping structure’ pour la segmentation hiérarchisée en motifs, phrases et sections, ‘metrical structure’ pour la description métrique sur différents niveaux, ‘time-span reduction’ pour hiérarchiser les hauteurs selon leur position métrique et ‘prolongation reduction’ pour l’expression des tensions, détentes et progressions harmoniques ou mélodiques. Une principale critique de cette théorie est que le parallélisme, c’est à dire les relations de similarité entre des séquences de plusieurs événements n’est pas étudié. Par ailleurs, le fait de choisir la partition comme niveau de départ place plutôt l’analyse du côté du compositeur alors qu’elle se revendique proche des intuitions d’un auditeur.

2.2.4 L’analyse paradigmatique

Si l’analyse paradigmatique reste imprégnée de l’idée propre à la linguistique qu’une séquence musicale peut être segmentée en différentes unités syntaxiques sur plusieurs niveaux hiérarchiques, deux de ses aspects marquent

une rupture par rapport aux théories précédentes :

- son renoncement à amorcer une analyse avec pour point de départ la structure attendue de l'œuvre ;
- la nouvelle place accordée aux répétitions par rapport aux groupements autour des accents sur lesquels se focalisaient les théories précédentes.

La recherche ‘sans a priori’ de principes fondamentaux

David Lidov [24] présente le problème de l'analyse de la façon suivante :

« The analysis of given texts must be our only source of theoretical knowledge about music. Therefore, we must find a technique which will analyse a whole text without relying on a prior separation of its parts, on any a priori normative assumptions about musical structure, or on recourse to undefined abstractions. »

L'analyse paradigmatique doit se faire à l'aide de principes fondamentaux qui seraient communs à tous, c'est à dire universels, de manière à assurer une objectivité scientifique, garante de vérité. Le respect de cette objectivité devient alors plus important même que la pertinence perceptive des résultats des analyses. Cependant, sans critère d'évaluation autre que l'objectivité, toutes les analyses deviendraient possibles. Ce problème sera plus largement repris dans l'une des sections suivantes intitulée 'La tripartition de Nattiez'.

L'analyse des répétitions

La répétition n'est pas un concept nouveau en analyse, mais l'analyse paradigmatique en a fait à la fois une notion de base, en a amorcé la formalisation et lui a accordé l'avantage sur les groupements accentuels pour établir les critères de segmentation.

La systématique introduite par Ruwet

Ruwet [46] propose une procédure quasi-algorithmique de segmentation d'une séquence musicale de haut en bas (des plus grandes unités vers les

plus petites). Il propose dans un premier temps de rechercher les plus longs segments identiques, puis d'examiner les autres pour chercher s'il ne sont pas transformations l'un de l'autre.

En proposant une telle méthode, Ruwet repose la question de la systématique dans l'analyse musicale. Faut-il que les critères de segmentation de l'analyste soient explicites ? Il semble que ce soit la direction prise par l'analyse paradigmatique, allant de pair avec la recherche de structures sans a priori sur leur forme. Il devient alors plus difficile de faire la part entre ce qui est objectif et ce qui ne l'est pas dans l'analyse. La recherche d'une systématique et le but avoué de ne considérer que des critères universels de segmentation partent d'une volonté positive, mais deviennent une charge contre la théorie lorsque la part de la subjectivité se retrouve cachée. Ainsi, la segmentation proposée par Ruwet n'est aucunement explicite car de nombreux termes restent flous, ni systématique car son application stricte à l'exemple qu'il donne aboutirait à un résultat bien différent de celui proposé. En effet, la procédure proposée, trop rigide pour être appliquée telle quelle à une œuvre musicale, n'est même pas respectée dans l'exemple donné en illustration : les plus longs segments identiques sont choisis plus qu'ils ne sont issus d'une procédure systématique, et d'autres segments plus longs qu'eux sont passés sous silence, sans doute pour que l'analyse corresponde mieux aux résultats attendus.

La tripartition de Nattiez

Nattiez [35], à partir des concepts établis par Ruwet, développe la théorie de la tripartition du phénomène musical. Le phénomène musical se divise en trois niveaux qui sont le niveau poïétique (ce qui est pensé par le compositeur) le niveau neutre (la partition objective) et le niveau esthésique (ce qui est perçu par l'auditeur). Le niveau neutre de cette tripartition est souvent remis en cause, car il explicite ce qui peut apparaître comme une faiblesse de l'analyse paradigmatique : la volonté d'objectivation. En effet, l'existence même d'un niveau neutre est-elle possible ? Le souci d'objectivité de l'ana-

lyse paradigmatique contient en lui-même une double limitation. D’une part il est très difficile voire impossible à définir sans faire référence à des notions non objectives du phénomène musical, et d’autre part il ne constitue pas en soi un critère suffisant de validation d’une analyse. En effet, on pourrait potentiellement effectuer une infinité d’analyses ‘objectives’ **donner exemple, ce qui leur enlèverait tout intérêt puisqu’elles seraient alors toutes également valides. En effet, comment adopter un critère qui privilégie une analyse objective par rapport aux autres sans trahir le principe même d’objectivité? Peut-être que ce principe correspond à un but que l’on peut viser sans devoir nécessairement l’atteindre.

Un critère récurrent de validation d’une analyse serait alors par exemple sa cohérence perceptive.

2.2.5 La perception, un critère récurrent en analyse

La perception peut être un angle d’approche intéressant pour l’analyse musicale, car elle permet de se situer du côté de l’écoute, passage obligatoire pour l’appréhension de toute musique.

La théorie de la gestalt

Certaines approches revendiquent l’utilisation de principes théoriques perceptifs tels que ceux décrits dans la théorie de la gestalt (la ‘bonne forme’), par ailleurs utilisée très fréquemment de façon non explicite par de nombreuses autres théories se disant même parfois non perceptives.

La théorie de la gestalt repose sur trois principes. D’abord, les objets étudiés sont vus comme un tout composé de différentes parties elles-même décomposables en plusieurs parties et ainsi de suite. Par exemple, une partition est constituée d’accords eux-même décomposables en notes. Ensuite, les propriétés du tout dépendent non seulement des propriétés de ses constituants, mais aussi des relations qu’ils entretiennent entre eux. Par exemple, une mélodie n’est pas uniquement un ensemble de notes, mais se caractérise aussi par leur ordre et leur rythme. Enfin, la perception ne peut être vue sous

l'angle de la statistique, mais comme une interaction dynamique entre l'auditeur et un stimulus. Ainsi, lire une partition est une expérience différente de celle de l'écoute d'une œuvre dans le temps, qui nous force en quelque sorte à une lecture strictement chronologique.

Narmour

Eugène Narmour [37] propose dans sa théorie un vocabulaire de cinq 'archétypes mélodiques' basés sur le concept d'implication-réalisation à partir desquels peut se construire une analyse de mélodies opérée note par note. L'analyse peut en théorie se faire suivant deux approches et donc aboutir à des résultats opposés ce que Narmour traduit par le concept d'ambiguïté. La première approche, de bas en haut (de note à note), repose sur des règles perceptives supposées universelles. La deuxième approche, de haut en bas, est basée sur des règles issues de nos expériences et nos cultures individuelles. Narmour n'a formalisé que la première approche, et laisse à l'analyste le soin de faire intervenir ses connaissances dans la deuxième.

Les archétypes sont issus de deux hypothèses de base (r1 et r2) qui associées aux concepts de 'closure' (frontière) et 'syntactic parametric scale' (niveau de définition de la similarité entre deux éléments) permettent de décider si un élément peut être identifié comme point d'implication ou non, et si oui laquelle.

Les hypothèses de base sont :

- r1 : la similarité entre éléments implique une attente d'éléments similaires ($A+A$ implique A) ;
- r2 : la différence entre éléments implique l'attente d'éléments différents ($A+B$ implique C).

Les différents archétypes (répétition, changement, symétrie, diade, monade) sont appliqués aux paramètres d'intervalles, registres et hauteurs en utilisant les principes gestaltistes de similarité, proximité et direction commune. Par exemple, si deux intervalles ascendants de un ton se suivent, la première hypothèse (r1) implique qu'un même intervalle de un ton suive.

Cette approche a l'avantage de ne pas partir d'un modèle de structure a priori, et pourrait donc conduire à la découverte de nouvelles structures. Par contre, les hypothèses de base paraissent trop rigides, et ne font pas forcément sens perceptivement, même si les principes utilisés pratiquement font appel à la Gestalt.

Meyer

Leonard Meyer [34] élabore sa théorie à partir des concepts perceptifs d'implication et réalisation qu'il applique à l'auditeur. Pour lui, chaque stimulus peut se réduire à une donnée binaire : elle fait ou ne fait pas sens pour nous. Si un stimulus fait sens, on s'attendra à ce qu'il se répète. C'est ce que Meyer appelle l'implication. La non réalisation de l'attente (ou frustration) est une information importante dans l'analyse car elle est porteuse de contraste. En même temps, une attente qui est satisfaite constitue aussi un fait notable puisqu'il modifie l'état de l'auditeur (et par là peut-être les autres attentes dont il était porteur). Le sens qu'à un stimulus pour l'auditeur dépend donc à la fois des frustrations et des réalisations qu'il entraîne chez lui. Par ailleurs, Meyer définit certains 'archétypes' tels que celui du 'vide-remplissage' (gap-fill en anglais) qui a été par la suite largement remis en cause, et qu'il applique à la surface musicale pour expliciter son organisation. Par exemple, le 'vide-remplissage' suppose qu'un intervalle assez grand d'une mélodie soit suivi par la succession des notes qui ont été omises, c'est à dire par les notes non jouées de l'intervalle, ce qui nous semble être une contrainte très forte sujette à de nombreux contre-exemples.

Plus généralement, la méthode d'analyse préconisée par Meyer est de reconnaître les attentes créées par les éléments de la partition pour ensuite les confronter à leurs réalisations ou frustrations, et par là mesurer l'écart entre les attentes de l'auditeur et la partition.

La principale critique faite à cette théorie ne concerne pas la démarche de partir de considérations perceptives d'autant plus complexes qu'elles sont basées sur la notion d'émotion, mais le fait d'explicitement la structure musicale

par des abstractions type flèches d'implication-réalisation qui ne font pas forcément sens musicalement. En fait, Meyer ne propose pas le découpage classique de la partition en hiérarchie de segments. Par ailleurs, les règles qu'il propose pour repérer les implications dans la partition sont finalement assez floues pour que la complexité initiale du problème d'analyse ne soit pas vraiment réduite.

2.3 Les modèles théoriques en analyse computationnelle

Il existe beaucoup d'applications possibles de l'informatique en analyse musicale. En ce qui concerne notre projet de recherche, deux applications sont particulièrement intéressantes : la validation ou l'extension d'anciennes théories, et la définition de nouvelles théories. Sans faire de présentation exhaustive, nous allons maintenant détailler certaines approches qui nous semblent intéressantes.

2.3.1 Temperley

David Temperley [49] propose une théorie d'analyse, à partir de fichiers MIDI, reposant sur un vocabulaire de règles issues pour une grande partie de la théorie de Lerdhal et Jackendoff. L'analyse est répartie en différents sous-problèmes : la recherche du rythme, la recherche de frontières de groupement à l'intérieur de monodies, la séparation des voix dans une polyphonie et enfin la recherche des paramètres harmoniques (tonalité, degrés, modulations). Chaque sous-problème se traduit par un ensemble de règles auxquelles on associe une pondération suivant qu'elles sont satisfaites ou non. Différentes analyses sont alors testées et évaluées d'après la somme des différentes pondérations. L'analyse retenue est celle totalisant un score maximal. L'algorithme de recherche du score maximal utilise les techniques de programmation dynamique.

Cette théorie marque une différence avec les précédentes par le fait qu'elle est entièrement formalisée. On peut voir dans l'explicitation des pondérations des règles de groupement une extension de la *Generative theory of tonal music* de Lerdhal et Jackendoff. Certaines critiques adressées à cette dernière théorie restent valable, comme la non considération du parallélisme, ou la prise en compte de règles discutables quant à leur généralité. Par ailleurs, l'approche ne résout pas le problème de la polyphonie et ne fait qu'aborder le problème de l'ambiguïté sans y apporter une formalisation. Par contre, l'utilisation de l'informatique résout certains problèmes comme la non explicitation des règles de groupement ou la difficulté d'effectuer manuellement une analyse sur plusieurs exemples.

On peut constater que le fait d'analyser un fichier MIDI et non plus une partition pose de nouveaux problèmes comme la recherche du rythme ou la notation des hauteurs dans la bonne tonalité. De fait, toutes les informations musicales non présentes dans un fichier MIDI doivent être explicitées par la théorie.

2.3.2 Conklin

Un peu de la même façon que Meyer, qui proposait un analyse basée sur un schéma d'attentes dont on pouvait mesurer la pertinence en le confrontant aux notes réelles de la partition, Conklin et Witten [9] adoptent la démarche d'une analyse computationnelle prédictive associant à une partition un graphe de valeurs d'entropies traduisant la probabilité d'occurrence de chaque note pour l'algorithme. Un ensemble de quelques partitions est utilisé comme base de test pour initialiser les règles de probabilité par un algorithme d'apprentissage.

Cette approche prédictive rejoint la démarche des modèles d'analyse par la synthèse. Ce type d'analyse n'explicite pas la structure de l'œuvre, mais cherche à en générer un modèle proche, dont l'écart avec le modèle existant constitue une mesure de sa pertinence. En supposant que cette approche aie des résultats convaincants, il nous faudrait accepter l'idée que la musique est

un phénomène régi par de simples lois de probabilités. Heureusement, nous n'en sommes pas là. Le modèle que proposent Conklin et Witten ne semble guère convaincant pour les auteurs eux-mêmes. De même, son application dans le domaine de la génération automatique de mélodies à partir d'une base d'exemples ne semble pas apporter beaucoup de résultats.

Nous pensons que l'un des problèmes de cet algorithme provient du niveau d'étude de l'analyse : le modèle est généré note par note, à partir des informations contenues dans les quelques éléments qui précèdent la note que l'on cherche à prédire, et sans qu'aucune des informations sur les groupements de notes ou répétitions de motifs ne soit utilisée. Du coup, de nombreuses informations échappent à l'analyse. Pour pallier ce défaut, les auteurs proposent d'intégrer au système plusieurs points de vue à partir de propriétés supposées connues sur la tonalité, l'harmonie, le rythme, la métrique etc. Nous pensons qu'en effet seule la considération de propriétés à un niveau plus élevé que la note peut fournir au système les informations qui lui manquent. Mais ces propriétés constituent déjà pour nous une analyse, et donc les supposer comme données d'entrée revient à considérer comme acquise toute une partie de l'analyse, ce qui réduit le niveau de généralité du système. Par contre, il pourrait être intéressant d'aller plus loin dans cette démarche pour déterminer si, pour un style donné, certains paramètres musicaux tels que le rythme ou la métrique pourraient être reliés de façon probabiliste à d'autres paramètres comme le rythme harmonique ou les enchaînements d'accords.

2.3.3 La GCTMS de Cambouropoulos

Cambouropoulos [6] propose une théorie générale de la structure musicale basée sur des principes cognitifs et logiques. Partant d'un fichier MIDI dont sont connus les onsets (positions en millisecondes des attaques des événements dans le temps), hauteurs, durées et dynamiques (sur une gradation de 0 à 127, la dynamique représente l'intensité perceptive des événements), l'analyse consiste essentiellement en deux étapes. D'une part sont développés individuellement des composants analytiques basés sur des problèmes

musicaux tels que un système de détection du mode musical, un système de segmentation, un système modélisant l'accentuation et la structure métrique, un système de détection et classification de motifs. D'autre part, sont développées des fonctions pour permettre à ces composants d'interagir entre eux. Le résultat final consiste à exhiber des motifs d'une partition ainsi qu'une métrique et un mode.

Cette approche par modules est intéressante car elle permet de décomposer un problème unique en plusieurs sous problèmes dont on s'attend à ce que la complexité soit moindre. Nous en détaillerons certains aspects dans la suite du document.

2.4 Synthèse

2.4.1 Quelle méthodologie en analyse musicale ?

La recherche de structures a priori

Parmi les différentes théories qui précèdent, nombreuses sont celles qui supposent a priori du type de structure musicale qu'elle vont se donner pour objectif de retrouver dans l'œuvre à analyser. L'analyse consiste alors en la recherche d'une figure remarquable, un peu à la manière de l'observateur, qui confronté à un ciel étoilé s'efforcera de distinguer des formes qui lui sont familières.

Est-ce bien le rôle de l'analyse ? Oui, si on considère l'analyse avant tout comme une interprétation de notre propre vision de l'œuvre. L'analyse peut alors asseoir et renforcer sa propre théorie, l'œuvre analysée devenant un support, un outil de rhétorique.

On peut toutefois se demander s'il n'est pas possible de minimiser la part accordée à l'a priori de façon à mieux adapter l'analyse à l'œuvre étudiée tout en favorisant la possibilité de découverte de nouvelles structures.

Inversement, l'a priori est un aspect indispensable à l'analyse puisque une analyse n'a d'intérêt que parce qu'elle se distingue des autres par un critère

que l'on énonce a fortiori a priori. Par exemple, la tentative paradigmatique d'atteindre une objectivité exemplaire, ou de manière opposée l'affirmation que l'analyse est subjective et diffère suivant chaque individu ne peuvent servir de principes de base pour élaborer une théorie intéressante puisque toutes les analyses possibles, calculables ou imaginables seraient alors également variable.

Plutôt que de présupposer d'une structure fixée à l'avance, on peut alors s'appuyer sur des a priori les plus généraux possible, c'est à dire sur des principes d'organisation temporelle tels qu'ils fassent sens pour le plus grand nombre d'entre nous.

Nous avons choisi de faire reposer ces principes sur des critères perceptifs, car quitte à devoir prendre parti a priori sur l'angle d'approche de l'œuvre musicale, nous préférons nous situer du côté de l'auditeur plutôt que par exemple du côté du compositeur (qui peut avoir cependant des principes communs avec l'auditeur) ou de tout autre acteur choisissant par exemple du point de vue des mathématiques ou de l'astrologie comme a priori. Par ailleurs, nous pensons que choisir un point de vue ne condamne pas l'utilisation de principes généraux. En effet nous pouvons supposer que nous partageons tous certains critères de perception de base, sans lesquels nous ne pourrions communiquer et comprendre nos impressions. Par exemple, nous sommes généralement tous d'accord pour dire qu'un objet est vert, ou qu'un son est fort. En fait, plus les objets sont simples et de bas niveau, plus nous pouvons espérer en avoir la même perception. De même, si les différentes formes de structures musicales ont une histoire et évoluent dans le temps, nous pouvons espérer qu'il demeure certains invariants dans la façon dont on les perçoit. Nous allons donc nous donner pour objectif de pousser le plus loin possible la limite à partir de laquelle les objets musicaux perçus ne seront plus les mêmes suivant chaque auditeur et suivant chaque style de musique. Pour cela, nous allons proposer certains mécanismes de notre perception qui font sens pour nous, comme par exemple la répétition, et nous serviront d'exemples pour stimuler les intuitions qui nous serviront de guide

dans la construction de notre modèle d'analyse.

L'influence des autres domaines

Nous avons vu que beaucoup d'analyses cherchent à appliquer à la musique des techniques provenant de domaines qui lui sont étrangers. Mis à part la problématique de l'a priori que nous venons d'évoquer, la question se pose de savoir dans quelle mesure les techniques issues d'un domaine non musical peuvent s'adapter à un problème spécifiquement musical.

Une attitude fréquente consiste à rattacher la musique à un autre domaine plutôt que de la considérer comme un domaine à part entière. Par exemple, les statistiques proposent des formules pour classifier et décrire l'organisation des suites de nombres. Il est alors assez tentant d'appliquer ces formules telles quelles à un problème musical en adoptant une représentation numérique des paramètres musicaux, ce que permet par exemple la norme MIDI. Or, une telle représentation ne fait pas forcément sens pour ces formules et il sera nécessaire avant leur éventuelle utilisation de trouver une représentation qui leur soit adaptée, ce qui constituera une grande partie du problème. Seule une étude détaillée des caractéristiques propres à la musique permettra d'aboutir à une telle représentation.

La musique est un domaine avec ses propres spécificités. Cette affirmation s'avère pour nous d'autant plus importante que le codage numérique rend maintenant possible la transmission et l'application directe de méthodes entre des domaines très différents, ce qui peut donner envie de faire l'économie de la recherche d'une représentation appropriée.

Par ailleurs, nous pensons qu'il n'est pas toujours possible d'aboutir à une représentation directement utilisable par des techniques extérieures au domaine musical, et dans ce cas il nous faudra inventer les procédures mêmes d'analyse.

Le matériau d'étude

Le choix du matériau d'étude d'une analyse musicale n'est pas une tâche anodine. Bien souvent, il influencera la manière ultérieure de réaliser l'analyse. Le choix du matériau d'étude peut se faire à deux niveaux différents :

- le niveau du corpus ;
- le niveau de la représentation musicale.

Le choix du corpus détermine le niveau de généralité de l'analyse ainsi que le type de complexité musicale pouvant être traité. Par exemple, les théories d'analyse de la musique tonale pourront se servir d'heuristiques propres à ce langage, mais ne pourront alors être appliquées à d'autres musiques. Un exemple de complexité musicale concerne la différence entre polyphonies et monophonies. Aucune des approches computationnelles qui ont été présentées précédemment ne peuvent analyser la musique polyphonique, alors que les théories pré-computationnelles ne faisaient pas de distinction formelle entre monophonies et polyphonies, puisqu'elles autorisaient un certain empirisme pour l'extraction des passages mélodiques à analyser. On peut en déduire que l'analyse de musiques polyphoniques présente une complexité difficile à formuler par l'informatique. Nous proposerons une solution à ce problème dans les chapitres suivants. Par ailleurs, la généralité du corpus sera l'un des objectifs que nous nous fixerons dans l'élaboration de notre modèle d'analyse.

Le choix de la représentation musicale a lui aussi une influence dans l'élaboration des différentes étapes d'analyse. Souvent, ce choix se fait en fonction des contraintes liées au support d'étude, suivant que les musiques à analyser sont disponibles au format MIDI, audio ou sous forme de partition. Dans ce cas, les méthodes d'analyse s'adaptent aux formats de représentation disponibles. Avant l'utilisation de l'informatique, la partition était souvent prise comme support de base. Certaines informations manquantes comme la répartition des voix, les ornements ou les groupements cadentiels pouvaient être rajoutés manuellement suivant les besoins et étaient considérés comme allant de soi. Cette façon de procéder est remise en cause lorsque l'analyse

est considérée sous l'angle de la perception et lorsque ses procédures doivent être formalisées. Si l'analyse doit se faire suivant des principes perceptifs, la représentation par la partition reste-elle légitime ? En effet, les rythmes et les notes inscrits dans la partition ne correspondent pas forcément à ce qui est perçu par l'auditeur. Mais alors, de quel niveau partir ? Celui du signal ? Il n'est pas certain du tout que nous percevions avec plus d'acuité les signaux que les hauteurs du tempérament. De fait, cette question ne semble pas avoir de réponses précises, et plusieurs théories computationnelles adoptent une représentation générale permettant de représenter plusieurs informations symboliques en fonction de leurs besoins, comme par exemple la norme MIDI qui permet de coder le rythme, les hauteurs, les dynamiques, la répartition des voix etc.

Nous pensons que la question du matériau d'étude peut alors être reliée à celle de l'ambiguïté musicale qui selon nous diffère selon les niveaux auxquels on se place dans le flux musical.

L'ambiguïté de l'analyse

L'ambiguïté n'est pas une notion proprement musicale, mais très générale. Par exemple, dans le domaine de l'image, des structures graphiques peuvent être ambiguës lorsqu'elles sont explicables de différentes façons¹. Dans le domaine musical, l'ambiguïté de structure est plus difficile à déceler et reste l'un des principaux problèmes auxquels se heurte l'analyse. Nous pensons qu'une façon d'approcher cette ambiguïté est de considérer la manière dont elle varie suivant les niveaux d'analyse auxquels on se place. Une hypothèse que nous formons est que cette ambiguïté respecte une certaine continuité lorsque l'on passe du microscopique au macroscopique du flux musical, en allant d'un niveau de faible complexité vers un niveau de complexité plus grand. Nous pensons qu'il existe en particulier une étape à partir de laquelle l'analyse devient sujette à différentes interprétations possible et non plus reconnaissance non ambiguë de formes. Nous passons de l'analyse par reconnaissance à

¹voir notamment les exemples proposés par Rispoli [42]

l'analyse par interprétation. Cette étape se situe à l'endroit où dans le matériau musical les éléments ne font plus l'objet d'une description unanimement reconnue mais sont sujets à différentes interprétation. Dans le cadre de notre étude, nous situons cet endroit après la description des hauteurs, mais avant celle du rythme. Après la description des temps d'attaque des notes, mais avant celle des harmonies. Lorsqu'un accord est joué au piano, une oreille entraînée reconnaîtra, sans qu'il y ait discussion possible, les différentes notes le composant. Les différentes hauteurs ne seront pas toujours égales, selon le tempérament ou l'instrument choisis, mais ce qui compte est l'unanimité dans leur interprétation. Au contraire, la reconnaissance du rythme peut souvent différer selon chacun, sans que l'on puisse forcément trancher entre ces différentes interprétations. Si la partition contient une représentation possible du rythme, elle ne correspond pas forcément à ce qui sera entendu par la majorité des auditeurs, contrairement aux hauteurs. En amenant l'analyse sur la recherche de structures de plus haut niveau, nous nous rendons compte que les possibilités d'interprétation sont encore plus nombreuses. Par exemple, contrairement au rythme, la recherche de motifs ne possède pas toujours de solution privilégiée notée sur la partition. C'est que plusieurs motifs différents peuvent être contenus dans une même séquence, selon le point de vue adopté. Plus on monte dans le niveau d'analyse, plus les interprétations sont multiples.

Cette hypothèse nous pousse alors à considérer l'analyse de bas en haut, c'est à dire que l'on commencera par définir un niveau initial de représentation du flux musical pour l'élever progressivement par l'analyse.

2.4.2 Deux opérations de base pour la recherche de structures

Au regard des nombreuses théories ayant jalonné l'histoire de l'analyse musicale, les outils mis en œuvre pour manipuler et explorer le matériau musical sont finalement très peu nombreux, et reposent presque tous sur deux opérations de base très générales à savoir la segmentation et la comparaison.

L'opération de segmentation consiste à découper un objet en plusieurs parties (ou segments). Pour cela, il faut établir des frontières (ou délimitations, séparations, points d'articulations) sur l'objet étudié.

L'opération de comparaison permet d'établir des similitudes, différences, variations et plus généralement des fonctions entre les différentes parties de l'objet étudié.

Ces deux opérations se complètent et interagissent ensemble sur différentes échelles. En effet, pour segmenter un objet, il est nécessaire de comparer ses éléments. Mais pour déterminer les éléments à comparer, il faut les délimiter par une segmentation de l'objet.

La notion de segmentation

La segmentation d'une œuvre musicale en ses composants principaux peut constituer le résultat d'une analyse. Pour parvenir à ce résultat, d'autres segmentations intermédiaires sont souvent nécessaires.

On peut se demander de façon plus générale pourquoi cette opération est tellement courante dans nos activités de manipulation et d'analyse des objets qui nous entourent. Par exemple, est-ce qu'un ordinateur a lui aussi forcément besoin d'effectuer des segmentations pour manipuler et mémoriser des objets ? Est-ce que la notion de segmentation est propre à l'homme ou propre au monde qui nous entoure ?

En général, on segmente un objet pour qu'il nous apporte un supplément de sens. On effectue en quelque sorte un formatage de sorte que les différentes parties de l'objet soient plus faciles à appréhender que le tout. Plusieurs facteurs peuvent participer au gain de sens. Toujours, la notion de comparaison intervient :

- les différentes parties peuvent être de taille ou de forme plus facile à mémoriser ou imaginer, c'est à dire de tailles comparables ;
- les frontières des parties peuvent être perceptivement plus saillantes.

Dans ce cas, nous comparons des instants temporels. Par exemple, des

temps forts ou des accents peuvent induire la perception de changements, de débuts ou de fins ;

- les parties peuvent être similaires ou différentes. La comparaison se fait entre des séquences temporelles ;
- les différentes parties peuvent entretenir des rapports fonctionnels particulièrement perceptibles. Par exemple, on peut relier deux phrases musicales par la notion d’antécédent-conséquent, ou deux accords par la notion de tonique-dominante. La comparaison se fait alors soit entre des instants temporels soit entre des séquences temporelles.

La notion de comparaison

Lorsque deux objets nous sont présentés, il nous est difficile de ne pas les comparer. C’est une opération intuitive, presque inconsciente. Pendant l’écoute d’une pièce musicale, nous effectuons en permanence des comparaisons entre ce qui s’est produit et ce qui se produit. Par exemple, un accent n’est souvent perçu comme tel que parce qu’il diffère par rapport aux notes précédemment entendues. Un passage mélodique sera reconnu, et non simplement perçu, parce qu’il est déjà entendu.

La comparaison peut être évidente lorsque les objets sont très différents, ou identiques, mais elle s’avère plus délicate lorsqu’intervient la notion de similarité. A ce moment là, des critères différents selon les individus et les situations peuvent intervenir dans le jugement, et la comparaison devient une tâche très complexe.

Si l’activité de comparaison est reconnue comme nécessaire pour l’analyse musicale, alors l’utilisation de l’informatique implique sa modélisation. Nous proposerons quelques solutions à ce problème central dans les chapitres suivants.

2.4.3 L’informatique est-elle nécessaire à l’analyse musicale ?

Nous allons d’abord nous interroger sur le rôle que l’informatique peut jouer par rapport à la place de plus en plus grande accordée au systématisme en analyse musicale, puis nous présenterons de façon générale les avantages et inconvénients liés à l’utilisation de l’informatique en analyse. Enfin nous présenterons la façon dont nous avons décidé d’intégrer l’informatique à notre modèle d’analyse.

Une formalisation de plus en plus systématique de la musique

Comme le remarque M.Malt [27] dans les prolégomènes de sa thèse :

« La formalisation de la pensée compositionnelle, et son étude de façon systématique ne date que de la deuxième moitié du XXème siècle. »

Cependant, l’activité de formalisation non systématique a commencé bien avant, comme en témoignent les écrits de Rameau [39], pour devenir progressivement un domaine à part entière désigné sous l’appellation ‘analyse musicale’. Au passage, l’analyse musicale ne se réduit pas aux seules approches formelles, puisqu’il existe par exemple un courant d’analyse empiriste initié par Tovey [50].

La formalisation va toujours en s’accroissant et touche presque tous les domaines de l’analyse jusqu’à les faire se confondre parfois. Par exemple, la formalisation d’une pensée compositionnelle peut être prise comme théorie analytique et être alors appliquée à n’importe quelle musique, même celle ne relevant pas de cette pensée. Ce serait le cas si l’on analysait une musique de Chopin sous l’angle du sérialisme.

Cette évolution de la formalisation s’associe alors souvent à la recherche d’une plus grande systématité dans l’analyse, que l’on peut qualifier de systématité théorique, passant d’une description de la musique à travers des écrits à des théories de plus en plus formalisées, puis à l’élaboration de

modèles systématiques dont la systématique auparavant théorique est aujourd'hui vérifiable par une implémentation informatique. Parmi les récentes théories, nombreuses sont celles qui se veulent le plus générales possible. Il est par exemple courant maintenant de critiquer une théorie comme celle de Lerdhal et Jackendoff parce qu'elle ne s'applique qu'à la musique tonale, et donc n'est pas générale, de la même façon que Lerdhal et Jackendoff [22] présentaient leur *Generative Theory of Tonal Music* comme pouvant rendre compte de la totalité des intuitions de l'auditeur, généralité qui manquait aux autres théories de l'époque.

Face à la systématique théorique qui provient de notre désir de définir une théorie applicable à toutes les musiques se présente maintenant la systématique computationnelle, ou l'automatisme, c'est à dire l'implémentation informatique totale ou partielle d'une théorie systématique qui, lorsqu'elle est sincèrement appliquée, c'est à dire sans que soient ajoutées des modifications propres à chaque musique étudiée, permet d'effectuer une analyse en suivant automatiquement les commandes d'un programme. Ceci n'est d'ailleurs pas sans dangers puisque par la même occasion s'augmente alors considérablement le risque de trouver des contre-exemples à la théorie et donc de contredire sa systématique théorique. En effet, beaucoup de théories non formulées par l'informatique mais se voulant générales s'autorisent une modification des règles d'analyse selon les cas de figure rencontrés ce qui limite un peu artificiellement le nombre de contre-exemples à la théorie. De telles modifications ne pourraient être apportées à un programme sans que son schéma général d'implémentation et du coup la théorie dans son ensemble ne soient modifiés. La systématique (ou automatisme) computationnelle devient alors un moyen efficace de validation d'une théorie qui se réclame systématique.

Avantages et inconvénients de l'analyse computationnelle

Le principal avantage de l'informatique est son infaillibilité. Sans elle, la systématique computationnelle aurait peu d'intérêts. L'infaillibilité fait aussi de l'informatique un support idéal pour l'expérimentation et l'implémenta-

tion de modèles qui, en cas de mauvais résultats, seront seuls à pouvoir être remis en cause.

Mis à part l'infailibilité, l'un des atouts majeur de l'informatique est sa puissance de calcul. Grâce à cette puissance, des calculs antérieurement très longs peuvent maintenant être réalisés en un court laps de temps, ce qui permet d'en faire plus et donc de tester un plus grand nombre d'hypothèses.

Ces deux critères ont permis le développement de nouvelles procédures d'analyse et l'émergence de nouvelles théories impossibles à formaliser dans le détail jusque là.

Par contre, ces critères restent à adapter à la musique en particulier pour reconnaître les objets qui nous apparaissent comme similaires alors que la machine les mémorise comme différents, ne faisant pas a priori de différence entre des valeurs proches ou extrêmement éloignées. De ce point de vue, la trop grande précision de la machine ne peut être directement utilisée en analyse musicale, et devra faire l'objet d'approximations, de manière à pouvoir rapprocher des objets musicaux physiquement différents qui devront cependant être interprétés comme étant les mêmes. Ce sera le cas de la plupart des composants musicaux comme par exemple les durées, des rythmes ou des motifs.

Pour quelle utilisation de l'informatique ?

Parmi les différentes théories que nous avons présenté, celles qui proposent les résultats les plus aboutis accordent une grande importance aux intuitions de l'analyste. Autrement dit, les résultats des théories computationnelles sont loin derrière ceux des théories traditionnelles. On pourrait alors se demander dans quelle mesure l'informatique peut contribuer à faire progresser l'analyse. Faut-il considérer l'ordinateur comme un simple outil, et si oui quelles tâches peut-il remplir ? On pourrait considérer l'outil informatique sous trois aspects différents :

- son autonomie ;
- son champ d'application ;

- la généralité/qualité de ses traitements.

Les théories computationnelles d'analyse musicale ont souvent pour idéal de concevoir un système automatique capable de découvrir toutes les structures contenues dans n'importe quel corpus, ce qui peut sembler une idée séduisante. Faut-il pour autant transformer cet idéal en objectif, et si oui comment ? Les théories les plus récentes montrent que le problème se pose toujours d'équilibrer les contraintes provenant d'objectifs ambitieux avec le niveau des résultats que l'on peut objectivement s'attendre à obtenir dans le temps limité de la recherche. Par exemple, si l'on articule le critère de généralité/qualité des traitements avec le champ d'application, c'est à dire pour notre domaine l'exhaustivité/finesse d'analyse avec la diversité des corpus pouvant être analysés, la recherche d'un modèle intégrant finement tous les aspects musicaux sera très complexe à mettre en œuvre, et devra se limiter à l'étude de musiques assez simples (monophoniques le plus souvent) provenant d'un répertoire réduit.

Inversement, un modèle pouvant s'appliquer à un large corpus de musique polyphonique ne pourra sans doute réaliser une analyse détaillée prenant en compte l'ensemble des interactions entre les événements musicaux, mais devra choisir ses priorités parmi les aspects musicaux à considérer.

Par ailleurs, le critère d'autonomie est souvent opposé à la généralité ou qualité des traitements. Pour qu'un système informatique fasse des traitements de qualité, il doit sacrifier à son autonomie, et doit être guidé par un utilisateur expert. C'est sans doute vrai dans le cas d'un projet où chaque utilisateur définit lui-même ses objectifs, que l'ordinateur ne saurait deviner. Par exemple, un utilisateur voudra examiner la possibilité d'une symétrie de forme entre différents préludes de Bach, pendant qu'un autre s'intéressera aux différentes transpositions d'un motif donné au sein d'une fugue. Cependant, parfois les objectifs peuvent être supposés communs aux différents utilisateurs, par exemple lorsqu'il s'agit d'évaluer la perception des composants structurels de base d'une musique.

Nous avons choisi dans cette thèse une approche progressive basée sur l'idée d'un système d'analyse général et automatique basé sur des critères perceptifs, tout en portant l'accent sur la qualité des traitements plutôt que sur leur généralité en adoptant l'hypothèse que certains aspects musicaux peuvent être considérés dans un premier temps comme indépendants. Par exemple, nous pensons que la notion de répétition est essentielle pour la détection de motifs. Nous en proposerons alors un modèle qui laissera de côté d'autres aspects musicaux tels que les phénomènes d'attente ou de tension-détente. Nous irons même plus loin en jugeant que certains événements musicaux peuvent ne pas être pris en compte, par exemple les notes qui ne sont pas situées sur les pulsations. Inversement, notre étude du rythme se fera sur la base d'une modélisation de l'attente, considérée plus importante que celle des répétitions. Cette simplification pourra progressivement aller dans le sens de l'intégration de l'ensemble des aspects ou événements musicaux dans l'analyse. Mais dans un premier temps, elle nous permettra d'appliquer notre modèle à un contexte musical riche, complexe et varié.

Par ailleurs, plutôt que d'expérimenter exhaustivement les dernières techniques issues des différents domaines scientifiques, nous pensons que le principal problème aujourd'hui réside dans la compréhension des phénomènes musicaux. La complexité ou la modernité des techniques que nous déciderons alors d'employer ne constituera pas une justification en soi, ces techniques devant être employées à la requête des problèmes musicaux eux-mêmes et non l'inverse.

2.5 Notre modèle d'analyse

Nous proposons une analyse automatique par étapes d'un flux musical polyphonique, de bas en haut, du niveau de la note vers les structures rythmiques puis motiviques. Nous espérons ainsi faire face à l'ambiguïté musicale en l'accompagnant progressivement dans sa complexité. Nous nous plaçons par ailleurs du côté de l'auditeur qui serait confronté à l'écoute d'une sé-

quence musicale symbolique type MIDI.

2.5.1 Le matériau de départ

Nous avons choisi de démarrer notre analyse sur la donnée d'un fichier MIDI dont les hauteurs, onsets et durées sont connus mais pas le rythme. Les hauteurs sont considérées comme faisant partie du tempérament égal. Nous nous situons ainsi à un niveau particulier délimitant l'analyse par reconnaissance de l'analyse par interprétation telles que nous les avons définies dans la section intitulée 'L'ambiguïté de l'analyse'. Par ailleurs, dans un premier temps, nous n'intégrerons pas dans notre modèle le paramètre de dynamique sauf de façon optionnelle pour l'extraction de la métrique. Nous pensons que la dynamique n'est pas le composant musical essentiel, aussi ce choix permet de simplifier les algorithmes sans que les conséquences ne soient trop importantes pour l'analyse. L'information sur la séparation des voix ne sera pas non plus considérée car ce paramètre reste ambigu lorsqu'on se place du côté de l'auditeur, et ne fait pas sens pour la musique générée par des instruments polyphoniques comme le piano, pour lesquels on ne peut pas dire que son extraction coule de source.

2.5.2 Le rythme, un composant de base

Nous considérons dans un premier temps le problème de l'extraction rythmique. Nous proposons d'extraire une pulsation, puis une métrique et enfin d'opérer une quantification de la séquence à analyser.

L'analyse de la pulsation joue un rôle important car c'est sur elle que nous nous baserons pour analyser le mètre, quantifier la séquence ou rechercher des motifs. Notre approche englobe une modélisation du phénomène d'attente associée à une hiérarchisation accentuelle de la séquence à l'aide de pondérations que nous appellerons marquages. Nous verrons ensuite que l'extraction de la métrique requiert un autre type de modélisation et un emploi différent des marquages. Au final, nous obtenons une segmentation de

la séquence de départ d'après notre extraction des positions de battues de pulsation et de barre de mesure. La séquence est ensuite quantifiée à l'aide de ces informations de manière à régulariser chaque intervalle de temps entre deux battues successives. La séquence quantifiée fournira alors le nouveau matériau d'analyse pour l'extraction de répétitions.

2.5.3 La recherche des répétitions motiviques

Dans un premier temps, nous définissons une mesure de similarité entre deux séquences MIDI supposées quantifiées. Ensuite, nous analysons à l'aide de cette mesure les répétitions contenues dans un fichier MIDI dont la pulsation a été préalablement extraite. Enfin, nous extrayons certaines répétitions qui sont perceptivement plus pertinentes au regard de règles que nous choisissons. Une autre application des mesures de similarité entre deux séquences sera proposée pour effectuer la recherche d'un motif donné dans une base de données musicales.

Chapitre 3

L'analyse rythmique : le cas de la pulsation

Ce chapitre traite du problème suivant :

A partir d'un fichier MIDI dont sont connues les hauteurs, onsets, durées et intensités, extraire et représenter automatiquement le rythme tel qu'il pourrait être perçu par une majorité d'auditeurs.

Cette formulation nous conduira à dégager plusieurs sous-problèmes tels que comment extraire un rythme d'un fichier MIDI, comment le représenter, comment faire en sorte qu'il corresponde le plus possible avec le rythme supposé perçu par un auditeur moyen ? Nous présenterons plus particulièrement dans ce chapitre le problème de l'extraction de la pulsation. Les autres aspects du rythme, tels que la mesure ou le processus de quantification seront traités au chapitre suivant. Mais la première question qui se pose est d'ordre plus général : qu'est-ce qu'un rythme ?

3.1 Qu'est-ce qu'un rythme ?

On ne saurait concevoir de musique sans rythme. Il peut y avoir des musiques sans mélodies, sans nuances, sans instruments, et même sans sons

(nous pensons par exemple aux *4'33* de John Cage, ou aux silences de certaines pièces pendant lesquels le chef d'orchestre continue à battre la mesure), mais jamais sans rythme. C'est donc un élément essentiel à la musique. Mais comment le définir ?

Si l'analyse rythmique a déjà fait l'objet d'une littérature abondante, ce domaine reste toutefois difficile à cerner du fait des multiples définitions possibles du mot rythme. Ainsi, Natiez [36] introduit le sujet ¹ en remarquant :

« Il n'est pas d'ouvrage sur le rythme qui ne s'ouvre avec une lamentation rituelle sur le désordre des définitions du rythme et ne donne, à l'appui, un florilège de ces définitions. »

(notez la récursivité de ce paragraphe)

Nous ne nous lancerons donc pas dans une nouvelle énumération, qui serait sans doute aussi sincère et rituelle que les précédentes, mais peu à même de nous faire avancer dans notre problème. Nous allons plutôt essayer de définir les invariants à l'origine de notre sensation du rythme sans nous enfermer dans la concision d'une simple définition qui ne pourrait à elle seule résumer toutes les acceptions possibles du mot. Par la suite, il nous faudra quand même donner quelques définitions plus précises qui nous permettront de décrire notre programme d'analyse, mais ces définitions n'auront pas de portée absolue. En effet, elles s'inscriront dans notre projet de développer un outil informatique d'analyse rythmique automatique. Nos définitions s'appliqueront à un certain corpus musical que nous essayerons de faire le plus large possible mais qui ne pourra recouvrir tous les rythmes possibles.

3.1.1 Un invariant : la temporalité

Nous avons vu que le rythme est un phénomène essentiel à la musique. Par contre, il ne lui est pas spécifique. On peut parler du rythme à trois temps du galop d'un cheval (voir figure 3.1) ou du rythme 'mécanique' du

¹pour se faire une idée de la diversité des points de vue, quelques définitions sont disponibles sur le site <http://www.chez.com/atrythme/fr/text003.htm>

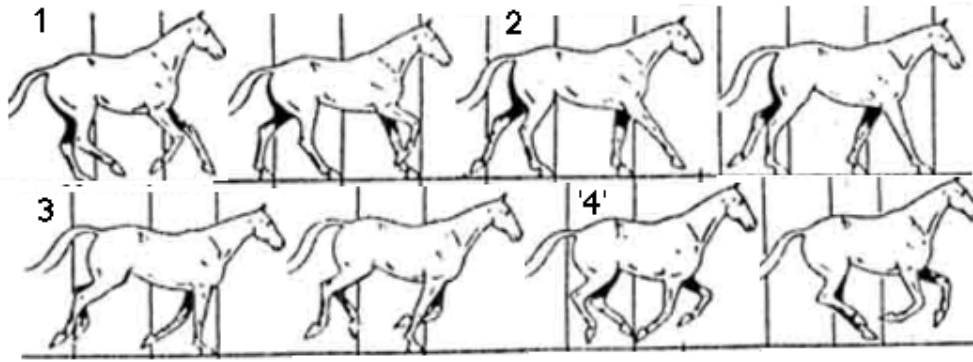


FIG. 3.1: La notion de rythme n'est pas propre à la musique. Ci-dessus, le rythme correspondant au galop du cheval décomposé en trois temps forts. Le quatrième temps, 'dans l'air', est rarement compté.

robot d'une chaîne d'assemblage.

Ces deux rythmes, même s'ils sont complexes et très variés (nous verrons toutefois par la suite que nous percevons moins les différences que les régularités), ne constituent pas de la musique. Peut-être faudrait-il dire que le rythme devient musical lorsqu'il est produit ou reproduit par l'homme, mais le terrain autour de cette question n'est pas très ferme, et y répondre sortirait des objectifs que nous nous sommes posés, donc nous ne nous y aventurerons pas.

Le rythme déborde donc de la simple question musicale.

En fait, l'élément qui donne sa particularité au rythme, c'est le temps 'physique'. Le rythme est une notion élaborée au fil de l'histoire pour nous permettre d'appréhender, mesurer et décrire la répartition dans le temps 'physique' des différents éléments d'un flux. Nous explicitons le rythme d'un flux temporel par la donnée de sa structure rythmique. Une structure rythmique se distingue des autres types de structure par le fait qu'elle se compose uniquement d'éléments temporels, c'est à dire d'éléments qui ont une durée et une position dans le temps.

Une autre acception du mot rythme s'applique parfois de façon plus générale à toute séquence d'évènements, le rythme étant alors le principe pré-

sidant à leur organisation. Ainsi, la *théorie du rythme* [26], proposée par Pierre Lusson, a pour objectif de décrire l'organisation de n'importe quelle séquence d'évènements, qu'ils soient musicaux, picturaux ou textuels. Un rythme pourra par exemple être issu d'une mélodie de hauteurs ou d'un vers de Plutarque.

De notre côté, nous limiterons le mot rythme à la structure proprement temporelle des évènements, c'est à dire résultant des relations qu'ils entretiennent dans le temps. Ce choix ne limite pas notre conception du rythme aux seuls onsets et durées. Par exemple, une mélodie de hauteurs pourra avoir une influence sur la structure rythmique, mais à condition que ses évènements soient associés à une réalisation temporelle.

3.1.2 Le rythme un phénomène immatériel ?

Avant d'analyser une structure rythmique, nous devons tout d'abord la 'localiser' : la structure rythmique est-elle contenue dans le flux (est-elle immanente au flux), ou est-elle imaginée par nous pour nous permettre d'appréhender ce flux ? En d'autres termes, faut-il chercher la structure dans le flux lui-même, ou dans les mécanismes par lesquels nous l'appréhendons ? Cette question rejoint le débat autour du niveau neutre tel que proposé par Natiez [35]. Existe-il une structure que l'on pourrait qualifier de neutre, c'est à dire indépendante de tout aspect extérieur à son support ? Si c'est le cas, nous serions bien incapables d'y accéder, puisqu'alors il nous faudrait oublier notre propre subjectivité ce qui est difficilement concevable.

Nous nous proposons alors de chercher le rythme en modélisant les conditions dans lesquelles on le perçoit. Pour percevoir un rythme, il faut lui donner une temporalité, et donc par exemple le jouer ou l'écouter dans le temps. A l'inverse d'une mélodie, un rythme se caractérise surtout par les relations temporelles qu'entretiennent ses éléments.

Il nous faut donc chercher le rythme en simulant son déroulement dans le temps.

3.1.3 Comment perçoit-on un rythme dans le temps ?

Pour établir un modèle d'extraction du rythme, nous voulons étudier les mécanismes cognitifs mis en jeu dans notre perception d'un flux temporel pour essayer de calquer le plus possible notre modèle sur eux. Ces mécanismes sont difficiles à saisir puisqu'ils ne nous sont pas directement accessibles. On pourrait par exemple chercher notre inspiration à travers quelques situations types dans lesquelles ils interviennent :

- ils peuvent être conscients ou inconscients : lorsque nous sommes occupés à une tâche, mais que le poste de radio est allumé, nous percevons le rythme sans le vouloir ;
- ils peuvent s'appliquer à différents supports de suites temporelles : lorsque nous visualisons un western (support visuel), ou que nous écoutons un prélude de Bach (support audio), nous percevons un rythme ;
- ils peuvent être mis en œuvre lors d'activités différentes : écoute, remémoration, visualisation ;
- ils peuvent changer suivant les connaissances investies par le sujet : les très appréciées dictées de rythme des cours de formation musicale nous apprennent à reconnaître des rythmes de plus en plus complexes.

Seulement, toutes ces situations peuvent difficilement aboutir à la construction d'un modèle, puisqu'elles sont à la fois trop générales et trop diversifiées. En fait, pour comprendre l'essence du rythme, plaçons-nous du côté de l'auditeur. Bien souvent, un hochement de tête, un léger balancement du corps ou un mouvement régulier du pied traduiront la perception d'une régularité dans le rythme. Cette régularité peut être ressentie par chacun d'entre nous. Un enfant de deux ans n'ayant pas de connaissances musicales se mettra lui aussi à marquer les temps d'une musique qu'il ne connaît pas du moment que le rythme est facilement perceptible. La généralité de ce phénomène donne toute son importance à l'analyse rythmique. Reste à définir le vocabulaire à employer pour en rendre compte. Pour cela, par facilité, nous allons nous appuyer sur le système de notation rythmique occidentale.

3.1.4 Nos représentations du rythme

S'il n'est pas commun à toutes les cultures, le système de représentation rythmique occidental est tout de même assez répandu.

Sur une partition rythmique, nous pouvons lire des noires, des croches etc. ainsi qu'une indication de tempo et des barres de mesure. Autrement dit, nous représentons l'unité rythmique de référence que l'on perçoit (par exemple tempo 60 à la noire, soit 60 noires dans une minute) par rapport à laquelle nous exprimons les autres valeurs rythmiques (les croches, les triolets, les pauses...). Par ailleurs, nous groupons l'unité rythmique de référence en mesures.

Définition. Un rythme se compose de :

- la pulsation ;
- la mesure ;
- les intervalles de temps proportionnels à la pulsation.

Seuls les rythmes répondant de cette définition nous intéresseront dans notre étude, même si d'autres rythmes, que nous pensons moins courants, pourraient être envisagés. Par exemple, on peut imaginer la situation suivante constituée d'un motif mélodique répété suivant plusieurs profils rythmiques différents à différentes vitesses : “jouer la séquence do-mi-sol trois fois de suite, la première fois en l'espace d'une seconde, la deuxième fois en deux secondes, et la troisième fois en trois secondes”. Chaque rythme de chaque répétition possèdera peut-être localement une pulsation, mais la séquence globale n'aura pas de pulsation ni de mesure.

Par ailleurs, en adoptant cette définition, nous supposons que le rythme esthétique (de la perception) peut être noté à l'aide du même langage que celui utilisé pour représenter le rythme poétique (de la partition), ce qui ne veut pas dire que les deux notations seront toujours les mêmes.

3.1.5 Conclusion

Chacun perçoit un rythme suivant ses propres processus cognitifs ce qui explique que selon les personnes, différents rythmes plausibles peuvent être perçus pour une même suite d'évènements. Cependant, bien souvent, une même unité de pulsation est perçue, et peut être représentée d'une même façon par un système de notation commun à tous. La notation permet alors d'unifier pour un même rythme joué les rythmes perçus par des personnes différentes. Notre objectif sera de proposer un algorithme permettant de représenter, dans un système de notation commun à tous, le rythme d'un fichier MIDI tel qu'il soit accepté par une majorité d'auditeurs. Pour cela, nous devons définir un langage de notation du rythme, ainsi qu'un algorithme permettant d'aboutir à cette notation à partir d'un fichier MIDI avec l'objectif de modéliser les mécanismes cognitifs de base qui nous amènent à la sensation du rythme.

Nous allons maintenant étudier séparément les trois composants du rythme que nous avons défini. Nous allons d'abord nous intéresser à l'extraction d'une pulsation à partir d'un fichier MIDI supposé en contenir une qui soit perceptible.

3.2 Extraire la pulsation

Dans cette section, nous allons d'abord préciser ce que nous entendons par le mot pulsation. Les différentes propriétés que nous aborderons seront présentées en ayant toujours à l'esprit l'objectif d'établir un modèle informatique d'extraction de la pulsation. Nous présenterons notre algorithme d'extraction automatique de la pulsation dans une deuxième partie.

3.2.1 Définitions

La pulsation est une régularité dans le temps qui nous sert de référence dans l'élaboration d'un rythme. Elle permet le découpage d'une séquence

musicale en plusieurs unités de même taille.

Elle est formulée mentalement lorsque nous sommes dans une situation d'écoute. On la désigne souvent par d'autres mots tels que tactum, pulse, période ou battue.

La pulsation, une référence rythmique naturelle

La pulsation peut être considérée comme la brique de base à partir de laquelle se constitue un rythme. Ses caractéristiques peuvent se comprendre par le fait que la dimension temporelle du rythme exige une mémorisation des événements passés pour pouvoir faire le lien avec les éléments présents et à venir. Autrement dit, il faut mémoriser une partie de la structure (le début) pour ensuite la comparer avec les éléments suivants. La proportion la plus simple à établir cognitivement étant l'égalité, il semble naturel de définir la pulsation, qui est une succession régulière d'intervalles égaux, comme unité de référence d'une structure rythmique.

Pour un rythme donné, il n'existe pas scientifiquement qu'une seule pulsation de référence possible, mais une infinité.

Musicalement, plusieurs pulsations peuvent aussi être imaginées. Une pulsation notée sur une partition ne sera pas forcément celle utilisée par l'interprète, qui lui-même ne percevra pas toujours la même que son auditeur, qui n'aura à son tour pas forcément la même pulsation de référence qu'un autre auditeur.

Les pulsations notées sur la partition sont les plus imprévisibles, donc les plus difficiles à retrouver, car elles dépendent du libre arbitre du compositeur. Parfois, la tâche est impossible à réaliser, soit parce que la notation est volontairement plus complexe que nécessaire et dans ce cas une pulsation plus simple sera perçue, soit parce que la notation ne respecte pas les critères perceptifs habituels, et dans ce cas la pulsation notée ne sera pas du tout perçue. Ces deux cas se rencontrent assez rarement pour les musiques que nous allons étudier.

Les pulsations de référence de l'interprète ou des auditeurs seront plus

facile à extraire. Elles entretiendront presque toujours un rapport de multiplicité simple, ce qui sera une bonne indication pour notre futur modèle. Notre objectif sera alors d'extraire l'une de ces pulsation, en essayant de se rapprocher le plus possible de la pulsation qui serait perçue par la majorité des auditeurs.

La pulsation, une régularité sans discontinuités

Il est très difficile de concevoir un rythme dont la pulsation de référence soit discontinue (exception faite des petites variations locales telles que le rubato ou le swing, que nous aborderons dans la partie suivante). Nous pensons que les raisons proviennent encore une fois de nos mécanismes perceptifs et cognitifs qui entrent en jeu à l'écoute d'un rythme. Nous pouvons essayer d'illustrer cela en comparant nos mécanismes avec les performances offertes par un ordinateur, ce qui rendra compte par la même occasion des problèmes pouvant être rencontrés pour établir un modèle informatique calqué sur les mécanismes en question.

Le rythme suivant serait très difficile à interpréter par nous sans la donnée d'une pulsation régulière de référence. Par exemple, la figure 3.2 montre une séquence de notes espacées de respectivement 1000, 501, 602, 776 et 1334 millisecondes. Pour pouvoir interpréter une telle séquence, il nous faudrait déjà connaître les rapports qu'entretiennent les durées entre elles. Pour cet exemple, on pourrait approximer ces rapports à respectivement 1, $1/2$, $3/5$, $7/9$ et $4/3$ de secondes. Même dans ce cas, la séquence resterait difficile à exécuter, puisque les rapports n'ont pas de dénominateur commun assez petit pour que leur comparaison fasse sens pour nous.

Un ordinateur n'aurait aucune difficulté à jouer l'exemple puisque son unité temporelle de référence est une période très petite (les logiciels que nous utilisons peuvent bien sûr jouer un rythme à la milliseconde près). Toute durée multiple de cette période peut donc aussi être jouée par l'ordinateur sans difficulté. Mais la période de référence la plus courante pour nous se situe autour de la seconde, qui représente par exemple le temps moyen nécessaire à

l'exécution d'un geste. La pulsation d'une séquence musicale, qui doit pouvoir être perçue, mémorisée et reproduite, est elle aussi proche le plus souvent de la seconde. Aussi, nous aurons beaucoup de difficultés à interpréter la suite de durées donnée en exemple car il nous faudrait percevoir les différents éléments selon une unité égale à $1/2 \cdot 5 \cdot 7 \cdot 3 = 1/210$ secondes. Cette différence de perception machine/être humain est une source d'enrichissement pour les compositeurs qui peuvent créer et faire jouer des rythmes aux ordinateurs en utilisant des périodes de référence très éloignées de celles qui nous sont familières.

Par contre, dans le cadre de notre modèle informatique, cette différence devient un véritable problème. En effet, il nous faudra établir des résultats conformes à nos périodes de références, et donc modéliser artificiellement notre système de 'constantes temporelles perceptives'. Par exemple, un système proposant un tempo de 500 à la noire ne sera pas très convainquant, même si la période correspondante est effectivement présente dans le fichier MIDI analysé.

Par ailleurs, une pulsation choisie arbitrairement en fonction des seuls premiers éléments d'une séquence rythmique ne conviendra pas forcément pour tous les événements de la séquence. Par exemple, la figure 3.3 montre une représentation rythmique de la suite de durées 1, $1/2$, $3/5$, $7/9$ et $4/3$ de secondes. On pourrait d'après les deux premiers éléments (1 et $1/2$ secondes) supposer que la pulsation de période $1/4$ de seconde conviendrait pour représenter le rythme, mais rapidement on aboutirait à une représentation très complexe des éléments suivants. La précision est encore trop grande et le choix que nous avons fait n'a pas de sens perceptif. Il faut alors rechercher une pulsation approximative telle que les rapports rythmiques entre chaque élément de la séquence soient représentés plus simplement, et par là deviennent plus compréhensibles pour nous. Nous allons maintenant détailler ce problème.

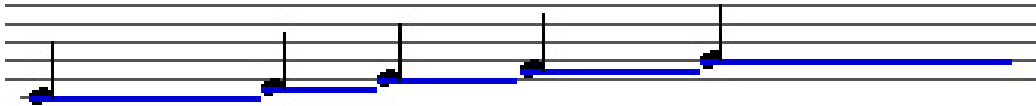


FIG. 3.2: Séquence de durées non quantifiées : 1000, 501, 602, 776 et 1334 millisecondes. Cette séquence est impossible à interpréter sans pulsation de référence.

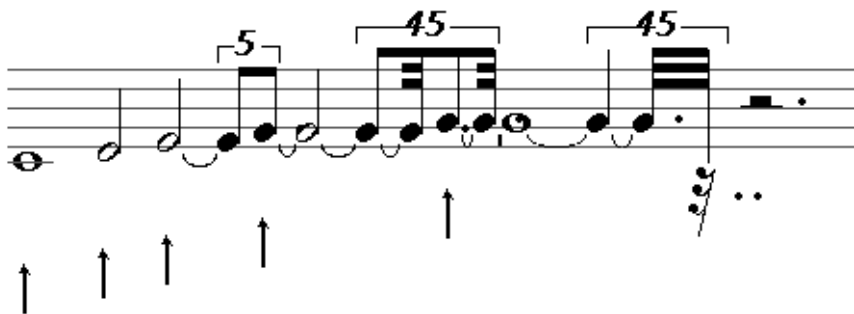


FIG. 3.3: Séquence de durées quantifiées : 1000, 500, 600, 777 et 1333 millisecondes (soit 4, 2, $12/5$, $28/9$ et $16/3$ multipliés par $1/4$ de seconde). Cette séquence est difficile à interpréter puisque la notation rythmique la plus exacte qu'on peut lui associer devient progressivement très complexe.



FIG. 3.4: Séquence de durées quantifiée : 888, 444, 571, 762 et 1333 millisecondes (soit $2/3$, $1/3$, $3/7$, $4/7$ et 1 multipliés par $4/3$ de seconde). Cette séquence est interprétable grâce à la relative simplicité de la notation rythmique. On peut alors penser que le rythme représenté peut aussi être perçu comme tel.

La pulsation, une régularité variable

Lorsqu'une séquence musicale est interprétée ou perçue par un auditeur, la pulsation peut varier dans le temps sans que l'on en soit conscient.

Par exemple, pour un interprète, la partition est découpée en unités de taille strictement exacte. L'interprète les perçoit alors bien comme exactes, mais il ne les jouera sans doute pas de façon strictement régulière. L'auditeur, lui aussi, perçoit des régularités dans les unités, mais cela ne veut pas dire que ces unités soient réellement exactes (on peut le mettre en évidence en enregistrant à la milliseconde près un auditeur battant la pulsation).

Pour modéliser ces variations, il nous faudra définir un seuil en deçà duquel les durées temporelles seront considérées comme les mêmes par l'algorithme. Une difficulté sera que le seuil peut changer suivant les personnes, et aussi suivant les contextes musicaux dans lesquels on le fait intervenir. Il nous faudra aussi trouver le meilleur compromis entre le taux d'erreur induit par les variations de la pulsation et la simplification de la notation rythmique résultante. Par exemple, la séquence précédente 1000, 501, 602, 776 et 1334 millisecondes peut être approximée par la séquence 888, 444, 571, 762 et 1333 millisecondes qui est représentée figure 3.4. Ce résultat a été obtenu par une approximation automatique des durées réalisée par le logiciel d'extraction de rythme que nous présenteront au cours de ce chapitre.

et de l'interprète. Elles sont le plus souvent continues, c'est à dire qu'elles se font graduellement, ou alors elles sont discontinues, et marquent un changement de pulsation, le plus souvent pour passer à une pulsation multiple de la pulsation de référence. Certaines variations portent un nom lorsqu'elles sont utilisées fréquemment et que leurs effets perceptifs sont très efficaces sur nous, par exemple rubato ou swing. Nous essaierons dans notre modèle de tenir compte des variations les plus courantes.

Les différentes composantes de la pulsation

Pour établir un modèle informatique d'extraction de la pulsation, il nous faut tenir compte des différents aspects que nous avons abordé dans les sections précédentes. Mais il est aussi nécessaire de définir plus précisément ce qu'on va rechercher et la manière dont on va le représenter. Pour cela, nous allons comme nous l'avons fait pour le rythme, utiliser la représentation de la pulsation dans la notation rythmique traditionnelle. Nous disposerons alors de tous les éléments pour entamer une recherche sur la modélisation de l'extraction proprement dite.

Définition. La pulsation, d'après le système de notation usuel, se décompose en une valeur que nous appellerons 'période', exprimée en nombre d'unités temporelles (par exemple la milliseconde), associée à une liste de positions temporelles que nous appellerons battues. D'autres termes sont parfois employés, comme fréquence, tempo ou pulsation pour désigner la période, et temps, phases ou pulsations pour désigner les battues, mais nous n'utiliserons dans la suite que les termes période et battue.

Il nous semble vraisemblable que ces deux éléments soient effectivement perçus lors de l'audition d'une séquence musicale, aussi nous allons maintenant essayer de modéliser un système capable de les extraire.

Comme nous l'avons déjà fait remarquer, nous n'essaierons pas forcément de revenir à la partition, dont le rythme peut être différent de celui qui est perçu. Nous évaluerons les résultats de notre système en fonction de leur

proximité avec la notation perçue.

3.2.2 Les différentes approches

Parmi les différentes approches qui se proposent d’extraire la pulsation, certaines simulent la notion d’attente. D’une certaine façon, cela revient à modéliser le passage du temps, à l’objectiver, en faire un objet que l’on va parcourir du début à la fin pour mesurer l’impact des évènements passés sur les futurs.

Par exemple, Large [20] et Scheirer [47] proposent de construire un modèle à partir d’oscillateurs non-linéaires sensibles à certaines fréquences qui réagissent à un signal audio pour prédire la prochaine pulsation.

L’algorithme utilisé par Scheirer présente l’avantage de s’appliquer directement au signal audio, contrairement à d’autres méthodes qui nécessitent la détection préalable des onsets. Par ailleurs, l’utilisation d’un banc de filtres réduit considérablement la taille en mémoire du signal ce qui facilite l’implémentation de l’algorithme en temps réel. Ce banc de filtre a par ailleurs une interprétation perceptive puisqu’il ne retient que les onsets qui seraient effectivement perçus. La séquence d’onsets est ensuite mise en entrée du banc d’oscillateurs qui entre en résonance avec les périodes les plus ‘résonnantes’. Une particularité de la méthode de Scheirer, tout comme celle de Large, est qu’elle ne prend en compte que certains des paramètres musicaux qui pourraient participer à notre perception de la pulsation. Si cette méthode fournit de bons résultats concernant certains styles musicaux (Country, Rock, Urban, Reggae etc.) où la pulsation est clairement énoncée par un instrument ou par une voix, elle devrait s’avérer moins fiable pour d’autres musiques. Par exemple, dans la musique pour piano, la pulsation résulte souvent de l’interaction des différentes voix et fait intervenir plusieurs paramètres tels que les hauteurs, les durées et les intensités.

Devant cet inconvénient, Scheirer propose d’utiliser son algorithme pour extraire une pulsation minimale, qui servira ensuite d’entrée à des algorithmes plus sophistiqués qui font intervenir les notions d’accents et de niveaux hié-

rarchiques, et qui détectent des pulsations de plus haut niveau. Seulement, nous pensons que les accents pourraient aussi être utiles pour la détection de la pulsation minimale.

Certaines approches utilisent alors la notion d’accent pour simplifier le problème, ce qui limite le champ d’étude et le nombre de valeurs possibles pour la pulsation. Par exemple, Temperley [49] analyse une musique tonale, dans laquelle les groupements de pulsations autres que deux ou trois sont interdits, avec une préférence pour les groupements par deux et une plage de valeurs de pulsations limitée allant de 400 à 1600 millisecondes. D’autres règles sont adoptées, par exemple pour favoriser les notes de durées longues, les accords de plusieurs éléments ou les pulsations les plus régulières possibles. Malgré cela, le nombre possible de solutions reste élevé, aussi le calcul se fait chronologiquement note par note, en choisissant pour chaque note une pulsation supposée la meilleure qui sera utilisée comme pulsation référence de la note dans la suite du calcul. Ce choix rend le modèle assez rigide, puisque une seule pulsation est associée pour chaque note et sa valeur ne pourra être remise en cause par les événements qui suivront.

Dixon [15] propose un modèle d’agents multiples non temps réel en deux temps : d’abord le modèle induit plusieurs pulsations, puis il cherche à les propager dans la séquence, c’est à dire qu’il recherche leurs battues. La méthode utilise une sorte de savoir musical basé sur la modélisation d’accents musicaux à partir de poids résultant de différentes propriétés musicales comme les hauteurs, durées, et positions temporelles (onsets). L’utilisation d’un savoir musical n’apparaît pas comme un inconvénient puisque les règles adoptées sont choisies pour leur généralité de façon à pouvoir les appliquer à plusieurs styles musicaux différents. Cette approche est particulièrement intéressante, et sera décrite plus en détail dans les section suivantes.

Un autre type d’approche consiste à utiliser l’apprentissage dans une étape précédant la recherche de pulsation proprement dite. Ainsi, Cemgil et Kappen [7] proposent une modélisation probabiliste des variations de tempo dont les paramètres sont calculés à partir d’une base d’exemples dont la pul-

sation est connue. Le problème est ramené à celui de l'optimisation de l'espace des solutions possibles. Comme l'indiquent les auteurs, cette approche est partiellement limitée puisqu'elle ne prend en compte que les informations sur les onsets. Par ailleurs, le fait même d'utiliser une base d'exemple réduit l'autonomie du système qui ne peut être confronté qu'à certains types de musiques cohérentes avec la base d'exemples.

Enfin, certaines méthodes analysent les répétitions de séquences rythmiques de façon à extraire des valeurs de périodicité. Par exemple, Desain et Siebe de Vos [12] proposent d'utiliser un système de fenêtrage : à chaque battue de pulsation supposée, une fenêtre est disposée sur les derniers événements de la séquence, puis des valeurs d'auto-corrélation sont calculées. L'autocorrélation est une mesure mathématique qui permet de rapprocher des séquences de poids ayant un même profil. Dans notre cas, les séquences de poids sont des suites d'inter-onsets. La distance entre deux suites d'inter-onset ayant même profil est considérée comme une période de pulsation. Cette période détermine la nouvelle valeur de déplacement de la fenêtre dans la séquence à partir de laquelle on calculera la prochaine période de pulsation. Cette méthode reste cependant limitée car l'autocorrélation ne peut extraire que les périodes de pulsation. D'autres approches doivent alors être envisagées pour la recherche des battues.

3.2.3 Un modèle d'accentuation des attentes

Il ressort des méthodes précédentes que les accents et l'attente jouent un rôle prépondérant dans notre perception de la pulsation, et leur modélisation fournit les résultats les plus satisfaisants. Nous allons donc essayer de conjuguer ces deux aspects dans notre modèle.

L'analyse des accents plutôt que celle des similarités

L'analyse basée sur les accents semble être une méthode plus utilisée que l'analyse basée sur les répétitions de motifs rythmiques. La recherche d'accents permet à la fois de prendre en compte d'autres paramètres musicaux

que les onsets et de situer avec précision les battues des pulsations. Mais il semble aussi que les accents jouent un rôle plus important dans notre perception d'une pulsation que les motifs rythmiques. Autrement dit, les événements situés entre les battues d'une pulsation interviennent plus souvent en tant que figures d'accents qu'en tant que profils rythmiques. Certaines musiques, comme le free-jazz, utilisent même presque exclusivement les accents comme éléments d'expression rythmique.

Nous allons donc nous attacher à modéliser les accents rythmiques. Pour cela, nous allons utiliser le concept de marquage. Le marquage d'une séquence d'événements a déjà été formalisé dans un cadre abstrait par Lussion [26] à travers sa théorie du rythme et par Mazolla [28]. La notion a été employée sans formalisme dans plusieurs études sur la musique (Cooper et Meyer [11], Lerdahl et Jackendoff [22], Cambouropoulos [6]). Cela consiste à d'abord choisir certaines propriétés musicales que l'on juge pertinentes au regard du problème qui nous intéresse, et ensuite à pondérer les événements d'une séquence musicale suivant qu'ils répondent ou non à ces propriétés. Par exemple, si nous choisissons la propriété 'contient plus de trois notes', les événements de plus de trois notes (les accords) pourront être pondérés 1 et les autres 0. Pour chaque propriété, différents poids peuvent être donnés. Il est alors intéressant de sommer ces poids pour obtenir une mesure de l'importance des différents événements au regard de l'ensemble des propriétés. Un événement qui possède de nombreuses propriétés aura un poids plus élevé qu'un autre qui n'en possède aucune. Nous procédons en quelque sorte à une hiérarchisation des événements suivant leurs caractéristiques par rapport à des propriétés que nous avons choisi.

Le choix de propriétés liées au phénomène d'accentuation rythmique nous permettra alors d'établir un profil d'accents sous forme de courbe de marquage.

La modélisation de l'attente

La modélisation de l'attente peut prendre en compte plusieurs paramètres. Parmi eux, le contexte culturel, c'est à dire les expériences passées, vécues ou apprises, joue certainement un rôle important. Seulement, il y a deux façon de le considérer : on peut le prendre comme acquis et à ce moment on le transmet sous forme de règles, un peu comme pour les systèmes experts, ou alors on suppose que l'algorithme peut l'apprendre et on l'entraîne alors sur différents morceaux que l'on juge représentatifs. Seulement, nous sommes alors renvoyés au problème de l'analyse préalable à l'apprentissage, problème qui nous préoccupe actuellement. Par ailleurs, définir un système de règles contredirait l'objectif de généralité que nous nous sommes fixé. Nous ne modéliserons donc pas le contexte culturel dans un premier temps.

Le deuxième paramètre concerne le contexte local, c'est à dire les événements récemment passés de la séquence étudiée. Ce contexte est plus facilement modélisable, car délimité et fini. Par ailleurs, il est nécessaire pour notre anticipation de la pulsation qui elle-même est nécessaire pour que l'on ressente les événements comme des battues au moment même de leur occurrence. Il doit y avoir coïncidence entre les battues, anticipées, et les événements qui leur correspondent, pour que l'on soit 'en phase' avec la pulsation. Un exemple de ce phénomène d'attente se produit souvent de manière perceptible lorsque la musique s'arrête brutalement. Bien souvent, on continue alors à marquer les battues, dans l'attente que quelque chose se passe.

Pour modéliser l'attente, nous pouvons nous inspirer du modèle proposé par Dixon, qui présente plusieurs avantages : il peut gérer l'attente de plusieurs pulsations, il peut prendre en compte d'autres paramètres musicaux que les onsets, et il peut s'appliquer aux représentations symboliques. Nous allons donc partir de ce modèle pour élaborer le notre, tout en essayant de pallier à certains de ses problèmes comme sa complexité exponentielle.

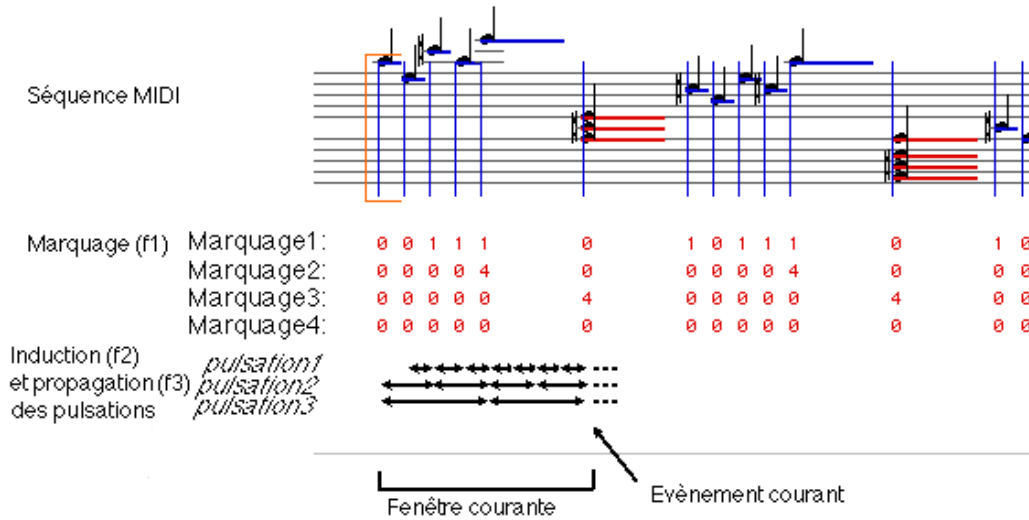


FIG. 3.6: La recherche des pulsation dans une séquence musicale MIDI. L'évènement courant de la séquence est marqué, puis des pulsations sont induites et propagées.

Description générale de l'algorithme

L'algorithme que nous proposons reprend de Dixon les notions d'accents et d'attente, mais nous lui apportons plusieurs améliorations que nous détaillerons à la fin de ce chapitre.

L'algorithme est constitué de trois principales fonctions (f1), (f2), et (f3), qui interagissent ensemble, plus une fonction facultative (f4).

Ces fonctions sont appelées à chaque déplacement d'une fenêtre qui parcourt l'ensemble de la séquence musicale évènement par évènement. Les évènements sont soit des notes, soit des accords du fichier MIDI analysé (un schéma général est donné figure 3.6).

Pour chaque nouvelle position de fenêtre :

- la fonction (f1) établit le marquage de l'élément courant. Ce marquage ainsi que ceux effectués sur les évènements précédents interviendront dans les décisions prises par les autres fonctions (f2), (f3) et (f4) ;
- la fonction (f2) induit les différentes pulsations possibles qui sont tenues

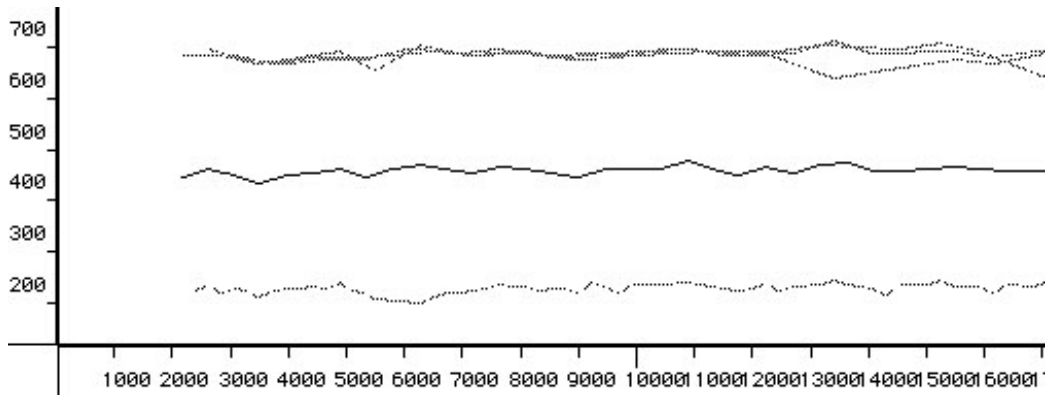


FIG. 3.7: Evolution temporelle de la Liste des Pulsations Possibles (l.p.p). Les valeurs de périodes (en millisecondes) sont représentées verticalement, et l'axe horizontal montre leur évolution dans le temps. Cinq courbes de périodes différentes sont montrées sur la figure. Trois se confondent presque en haut de la figure. Elles correspondent à la même période déphasée dans le temps. Les variations des courbes montrent de légers changements de tempo détectés par l'algorithme.

à jour dans une 'liste des pulsations possibles' (l.p.p). Chaque pulsation est constituée d'une liste de périodes, battues et poids de marquage, ces valeurs correspondant à la partie de la séquence déjà analysée ;

- la fonction (f3) détermine alors si l'évènement courant correspond ou non à la battue attendue d'une ou plusieurs pulsations de la liste. Chaque pulsation de la (l.p.p) est ainsi propagée dans toute la séquence (voir figures 3.7 et 3.9). La nouvelle valeur attendue de battue d'une pulsation est calculée en ajoutant la période courante à la dernière position de battue trouvée ;
- enfin, si l'utilisateur le souhaite, la fonction (f4) extrait de la (l.p.p) pour chaque position de fenêtre la pulsation la plus pertinente suivant nos critères.

3.2.4 Notre algorithme d'extraction

Au début de l'analyse, la liste des pulsations possibles (l.p.p) est vide. Ensuite, la séquence est analysée événement par événement. A chaque événement, les quatre fonctions (f1), (f2), (f3) et (f4) sont appelées, et la (l.p.p) est mise à jour à partir des sorties de (f1), (f2) et (f3). Suivant les cas, cette mise à jour consiste à :

- ajouter de nouvelles pulsations à la (l.p.p) à partir de la sortie de la fonction d'induction des pulsations (f2). A cette fin, un paramètre 'nb-puls-max' détermine le nombre maximal de pulsations pouvant être contenues dans la (l.p.p). 'nb-puls-max' est initialisé par défaut à 9 pulsations. Si le nombre maximal est atteint ou dépassé, les nouvelles pulsations qui pourraient être induites ne seront pas ajoutées à la liste ;
- supprimer de la l.p.p les pulsations dont la mise à jour n'a pas été faite depuis un certain temps. Un paramètre, 'contexte-pulse', détermine le nombre de battues consécutives qu'une pulsation peut avoir sans être mise à jour. Ce cas permet de traiter par exemple les pulsations dont la battue se fait sur les silences. Par défaut, 'contexte-pulse' est fixé à 8 battues ;
- mettre à jour les pulsations de la (l.p.p) dont la battue attendue correspond à celle de l'évènement courant (ce cas est déterminé par (f1) et (f3)). La période, les battues et les poids de ces pulsations seront alors mises à jour.

Les marquages utilisés (f1)

Les accents de pulsation nous induisent sur la bonne pulsation, comme le ferait un guide d'onde pour un signal optique. Lorsque nous nous écartons du droit chemin, les accents sont là pour nous recaler sur les temps forts correspondant à certaines des positions que nous avons appelé 'battues'. Pour déterminer la hiérarchie des accents, nous utilisons les marquages. Chaque paramètre intervenant dans la composition d'un accent peut ainsi être mo-

déliné séparément par un marquage spécifique. Nous avons choisi d'utiliser quatre marquages différents, ce qui ne constitue pas une liste exhaustive, mais un modèle de départ pour l'expérimentation. Chacun de ces marquages procure un poids pour chaque événement de la séquence étudiée, qui pourra ensuite être sommé aux poids des autres marquages de manière à obtenir un unique poids global pour chaque événement.

Les trois paramètres MIDI initialement disponibles (hauteurs, durées et onsets) ont tous été utilisés pour établir les quatre marquages suivants, deux paramètres (durées et onsets) interagissant dans la définition du quatrième marquage.

Le marquage 1 marque *les répétitions d'inter-onset* :

Nous définissons l'inter-onset comme étant l'intervalle de temps (en millisecondes) entre deux onsets successifs. Si comme pour la durée, l'inter-onset commence toujours sur un onset, il se termine forcément sur l'onset immédiatement successif alors que la durée peut se terminer à n'importe quel moment. Le marquage est booléen : sont marqués du poids 1 les événements dont l'inter-onset avec l'événement précédent est égal à l'inter-onset qui lui est immédiatement antérieur. Comme l'égalité de temps n'est pas absolu en musique, un seuil est pris en compte. Par défaut, la différence entre deux inter-onsets ne doit pas être supérieure à $2/10$ de leur moyenne. Ce marquage repère les régularités temporelles contenues dans la seule suite des onsets. Il traduit le fait que dans une séquence musicale, un onset régulièrement répété, c'est-à-dire répété à intervalles de temps égaux, est facilement perceptible par nous et nous conduit à préférer ces positions temporelles par rapport aux autres comme emplacement éventuel des battues de la pulsation.

Par exemple, la séquence d'onsets (en millisecondes) :

(0 10 50 90 110) est marquée

$$(* * 0 1 0)$$

Le marquage2 marque *les inter-onsets longs* :

Le marquage peut prendre plusieurs valeurs : sont marqués du poids nb les évènements dont l'inter-onset avec l'évènement suivant est plus grand que les nb inter-onsets consécutifs précédents. Un inter-onset est plus grand que le précédent si sa valeur lui est 1.2 fois supérieure. Ce marquage exprime le fait que les inter-onsets plus longs que les précédents sont susceptibles d'avoir leurs débuts situés sur une battue de pulsation. Perceptivement, cela correspond peut-être au fait qu'il est plus facile de mémoriser la longue taille de ces inter-onsets lorsqu'ils commencent ou finissent sur une position de battue. Nous n'avons pas marqué la fin d'inter-onset car il est sans doute plus difficile de mémoriser une taille d'inter-onset lorsque seule la fin se situe sur une position de battue, la position de début ayant alors déjà été entendue sans avoir été particulièrement mémorisée. Les fins d'inter-onsets ne doivent pas être confondues avec les fins de durées, qui par contre peuvent être à l'origine de la perception d'un accent lorsque par exemple plusieurs durées provenant d'onsets différents se terminent en même temps. Ce cas en effet ne se pose pas pour les inter-onsets qui par définition ne peuvent pas recouvrir les onsets. La prise en compte des fins de durées sera traitée par le marquage4.

Pour le marquage2, par exemple, la séquence d'onsets :

$(0 10 50 100 110)$ est marquée

$(0 1 2 0 *)$

Le marquage3 marque *les accords selon leur densité* (nombre de notes qu'ils contiennent) :

Sont marqués du poids 0 les silences ou accords de une note, de 1 les accords de 2 notes, et de 2 les autres accords. Les évènements

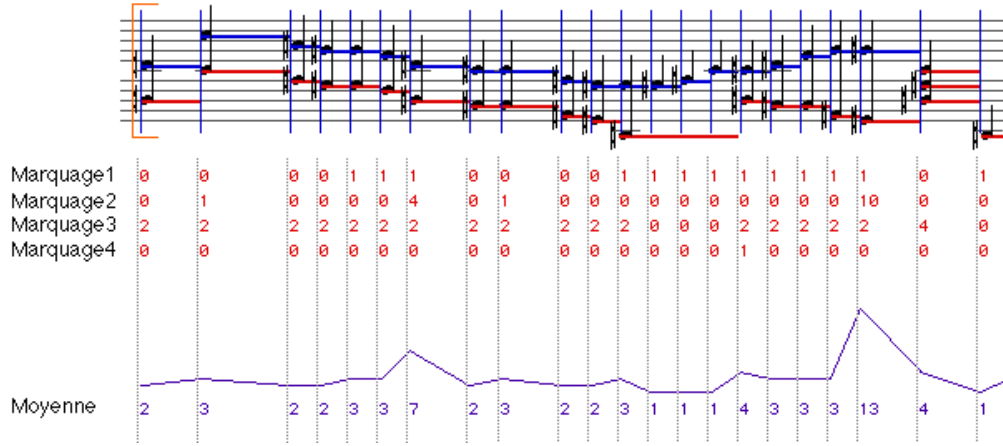


FIG. 3.8: Le marquage d'une séquence extraite de *the easy winners* de Scott Joplin. Les quatre marquages sont appliqués indépendamment puis linéairement combinés en une seule courbe (bas de la figure).

dont les onsets sont très proches sont préalablement regroupés en accords. Ce marquage exprime le fait que les évènements comportant plusieurs notes sont perceptivement plus accentués que les autres.

Par exemple, la séquence :

$$((C) (C E G) (D F) (G B D) (A)) \text{ est marquée}$$

$$(0 \ 4 \ 1 \ 4 \ 0)$$

Le marquage4 marque *les fins de durées* :

Sont marqués du poids nb-1 les évènements situés sur la position de fin de nb durées issues d'onsets différents. Une durée d1 issue d'un onset o1 est considérée se terminant sur un onset o2 si sa différence avec o2 - o1 n'excède pas 0.1 fois o2 - o1. Ce marquage exprime le fait que la synchronisation de plusieurs éléments simultanés d'une séquence est plus souvent réalisée sur les positions de battue que sur les autres positions.

Par exemple, la séquence (onset, durées) :

$$((0,90) (10,80) (50,40) (90,20) (110,*)) \text{ est marquée}$$

```

(((227 1 2227 2000 227 0 0 1) (228 6 2455 2227 227 1 0 1) (228 10 2909 2455 228 6 0 1) (228 11
((114 2 2341 2227 113 1 0 2) (114 7 2455 2341 114 2 0 2) (114 11 2909 2455 114 7 0 2) (114 12
((454 9 2909 2455 455 5 0 3) (454 10 3364 2909 454 9 0 3) (454 11 3818 3364 454 10 0 3))
((341 5 2455 2114 341 0 0 4) (322 6 3364 2455 341 5 0 4) (332 7 3705 3364 322 6 0 4))
((1591 1 3591 2000 1591 0 0 5))
((1250 1 3591 2341 1250 0 0 6))
|((682 1 3591 2909 682 0 0 7)
((228 1 3705 3477 228 0 0 9) (227 1 3931 3705 228 1 0 9)))

```

FIG. 3.9: Une partie de la liste des pulsations possibles correspondant à l'analyse de l'extrait montré figure 3.10. Chaque ligne contient la liste des différentes valeurs de période et battue d'une pulsation. Pour chaque élément d'une ligne (par exemple prenons le premier élément de la première ligne (227 1 2227 2000 227 0 0 1)) sont donnés respectivement : la période (en ms), le poids courant attribué à la pulsation, l'onset correspondant à la battue, l'onset de la battue précédente, la période précédente, le poids de l'onset précédent, un booléen pour la mise à jour, et un numéro d'identifiant unique pour la pulsation.

(0 0 0 2 0)

D'autres marquages sont envisageables, mais dans un premier temps nous nous sommes limités à ces quatre définitions qui constituent une base de départ pour expérimenter notre algorithme. Il sera toujours possible dans un deuxième temps de paramétrer ces choix, par exemple à l'aide d'une interface donnant la possibilité à l'utilisateur de définir ses propres marquages.

Une fois calculées, les quatre courbes de poids correspondant à chaque marquage seront ensuite normées puis sommées. Si certains marquages sont considérés comme plus importants que d'autres, ou si ils apparaissent moins importants mais devraient avoir autant d'importance, ils peuvent être pondérés plus fortement au moment de la sommation. Dans notre cas, le marquage3 (marquage des accords) est pondéré par le poids 2 de manière à étaler les poids de 0 à 4. Au final, l'application des quatre marquages à une séquence musicale aura donc pour sortie une seule suite de poids. L'exemple de la figure 3.8 montre l'application des quatre marquages à un extrait de *the easy winners* de Scott Joplin.

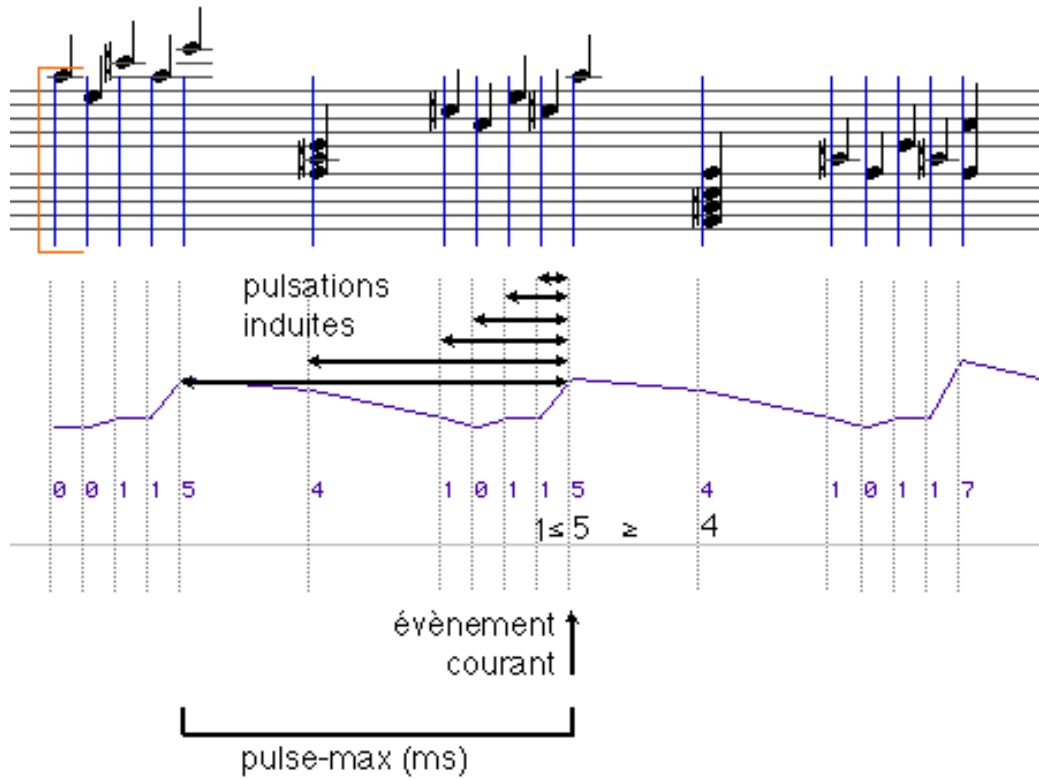


FIG. 3.10: L'induction de pulsations. Si le poids de l'évènement courant répond aux contraintes fixées, tous les inter-onsets entre cet évènement et les évènements précédents sont susceptibles d'induire de nouvelles pulsations, dans la limite de taille fixée par 'nb-puls-max' (par défaut fixé à 9).

L'induction de nouvelles pulsations (f2)

Cette fonction induit de nouvelles pulsations à partir d'une séquence de taille 'taille-contexte' à laquelle on a appliqué les marquages décrits dans la section précédente (fonction (f1)). Par défaut, 'taille-contexte' est fixé à 3000 ms. Pour qu'une nouvelle pulsation soit éventuellement induite, le marquage de l'onset du dernier évènement de la séquence doit avoir une valeur non nulle et supérieure ou égale à celle des évènements immédiatement précédents et suivants (voir figure 3.10).

Si cette contrainte est respectée, chaque inter-onset entre le dernier évènement et les autres évènements de la séquence est supposé comme période possible d'une nouvelle pulsation, sous réserve que le nombre maximal de pulsations 'nb-puls-max' n'ait pas été dépassé lors de l'itération précédente, et que la nouvelle période ne dépasse pas une certaine valeur 'pulse-max' (les périodes trop grandes sont éliminées). Par défaut, 'nb-puls-max' est fixé à 9 et 'pulse-max' est fixée à 2100 ms. La battue initiale de chaque nouvelle pulsation correspond alors à l'onset du dernier évènement de la séquence, et la fonction a comme résultat de sortie la liste des nouvelles pulsations.

La propagation des battues (f3)

La fonction détermine si l'onset courant correspond ou non à une nouvelle valeur de battue possible pour une ou plusieurs pulsations de la l.p.p.

Deux scénarii principaux sont à considérer :

- l'évènement courant correspond approximativement à la battue attendue d'une pulsation de la l.p.p. Pour cela, il faut que la période qui lui serait associée ne diffère pas de plus d'un certain pourcentage de la période attendue. Pour le vérifier, une fenêtre est définie, centrée sur la battue attendue de la pulsation.

De manière à prendre en compte les changements de tempo qui seraient perçus, la taille de la fenêtre est choisie proportionnelle à la période attendue de la pulsation. Le pour-

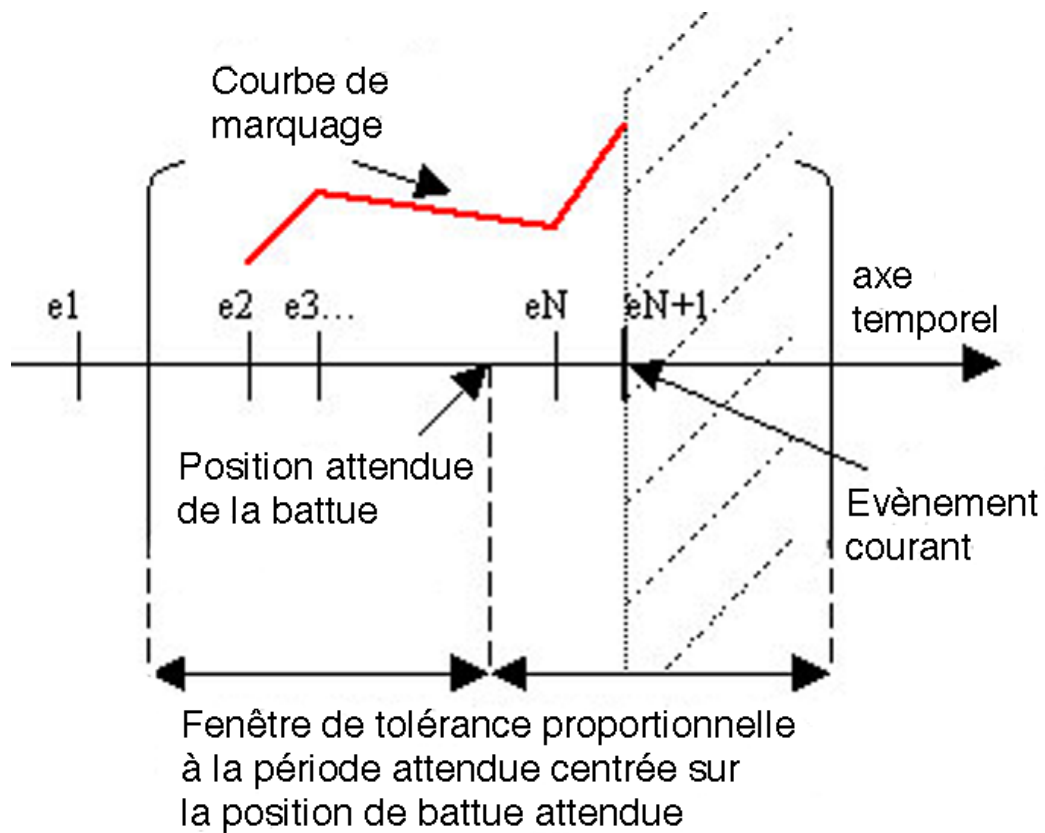


FIG. 3.11: La propagation des battues de pulsation. L'évènement e_{N+1} est confirmé par (f2) comme position de battue parce qu'il est situé dans la fenêtre d'attente de pulsation et que son poids est plus grand que celui de e_N , évènement précédemment choisi pour cette battue de pulsation.

centage de différence maximal autorisé entre la période attendue et la période approximée correspond au paramètre ‘taille-approx’ et détermine la taille de la fenêtre. Par défaut, ‘taille-approx’ est fixée à 0.2. Plus la période de la pulsation est grande, plus la fenêtre est grande et vice-versa. Ainsi, les grandes périodes peuvent varier davantage que les petites dont les changements nous sont plus immédiatement perceptibles. Si la position de l’évènement courant se trouve dans la fenêtre centrée sur la battue attendue, elle est acceptée comme nouvelle battue de pulsation. La nouvelle période de la pulsation est alors moyennée par l’ancienne période de manière à autoriser un changement progressif de tempo sans trop s’écarter de l’ancienne valeur en cas de variation ponctuelle non confirmée par les battues suivantes ;

- l’évènement courant répond au cas précédent, mais un autre évènement avait déjà été choisi comme nouvelle battue de la pulsation. Si l’évènement courant s’avère moins bien convenir que celui qui avait déjà été choisi (voir condition ci-dessous), rien ne se passe. Sinon, l’évènement qui avait été choisi comme nouvelle battue de la pulsation est effacé de la (l.p.p) et remplacé par l’évènement courant.

Pour déterminer si, pour une pulsation donnée, l’évènement courant convient mieux à la battue attendue que celui déjà choisi, on considère la fenêtre définie par ‘taille-approx’ et centrée sur la battue attendue de la pulsation. Un nouveau marquage est alors introduit de façon à ce que les évènements de la fenêtre les plus proches de la position de battue attendue aient un poids plus grand que les autres. Ce marquage est ajouté au marquage déjà établi pour les évènements de la fenêtre. Une fois ces opérations effectuées, l’évènement qui a le plus grand poids entre l’évènement courant et celui déjà sélectionné est choisi. La figure 3.11 illustre ce scénario.

La liste des pulsations possibles (l.p.p) est ensuite mise à jour comme nous l'avons décrit au début de la section intitulée 'Notre algorithme d'extraction'.

L'extraction de la meilleure pulsation (f4)

Cette fonction est facultative. A chaque itération de l'algorithme elle extrait de la l.p.p la pulsation jugée la meilleure. Cette étape n'est pas déterminante pour le fonctionnement global de l'algorithme, mais constitue une sortie intéressante pour l'utilisateur. Les meilleures pulsations de chaque étape sont mémorisées de façon à pouvoir être représentées globalement dans un graphe (voir figure 3.12). La détermination visuelle de la meilleure pulsation nous serait alors aisée avec un peu d'expérience : à chaque étape, nous essaierions de prolonger la courbe des anciennes valeurs de la façon la plus continue possible.

La réalisation automatique de cette opération s'avère plus délicate, car plusieurs points de vue différents peuvent intervenir selon les courbes rencontrées. Nous proposons une liste de différents critères, chacun permettant d'établir un classement des pulsations :

1. pour chaque pulsation de la l.p.p, multiplier le nombre de battues trouvées par la moyenne de la période. L'idée est de conserver les pulsations qui ont été propagées sur une grande partie de la séquence ;
2. pour chaque pulsation de la l.p.p, calculer la moyenne des poids de chaque battue. L'idée est de conserver les pulsations dont la battue est souvent placée sur des poids importants ;
3. pour chaque pulsation de la l.p.p, estimer la régularité du graphe d'évolution des périodes dans le temps. Ce critère favorise les pulsations qui restent stables dans le temps ;
4. pour chaque pulsation de la l.p.p, rechercher ses harmoniques parmi les autres pulsations de la l.p.p.

Deux pulsations sont harmoniques l'une de l'autre si leurs périodes ont un rapport de proportion simple et si leurs bat-

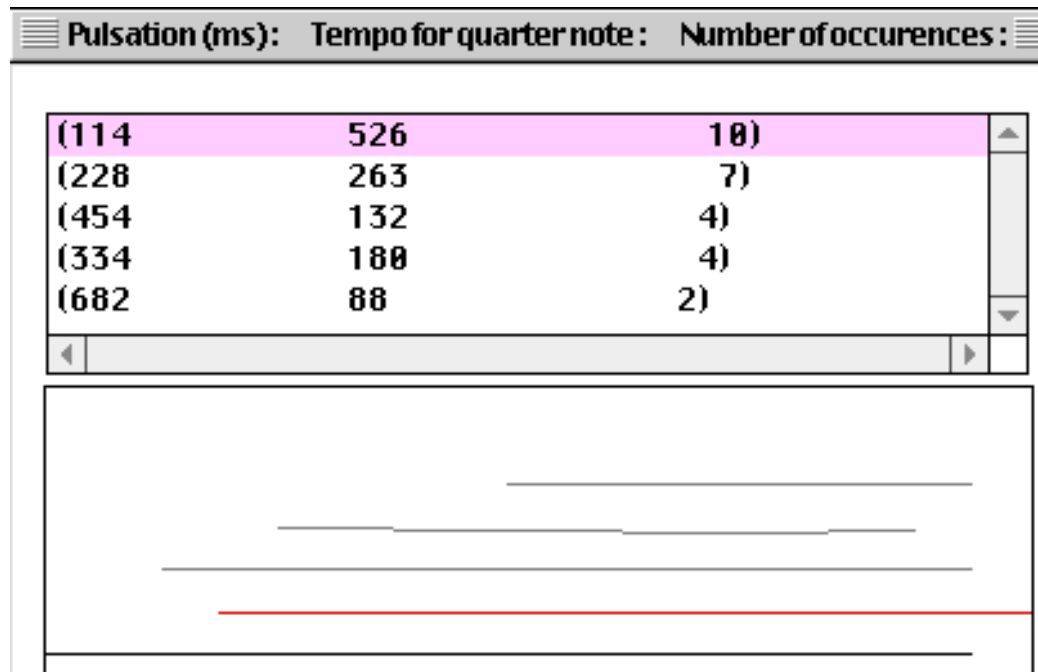


FIG. 3.12: La représentation graphique de la liste des pulsations possibles décrite figure 3.9. Les pulsations sont classées par ordre de préférence par la fonction (f4). La pulsation jugée la meilleure par l'algorithme est surlignée. Le graphique montre l'évolution dans le temps (axe horizontal) des périodes de chaque pulsation (axe vertical gradué en ms).

tues sont en phase. L'idée est d'extraire de la l.p.p un réseau de pulsations tel que les pulsations soient toutes liées par des rapports harmoniques simples. En effet, à l'écoute d'une séquence musicale, il est souvent possible de percevoir plusieurs pulsations multiples les unes des autres ;

5. mesurer la distance entre la fréquence moyenne de chaque pulsation de la l.p.p et la valeur 60 battues par minute. Les fréquences proches de 60 battues par minute sont très souvent utilisées sur les partitions et sont perçues avec d'avantage d'acuité que leurs valeurs multiples.

A la suite de différents tests, il est apparu que les critères 1 et 4 étaient les plus pertinents pour extraire une pulsation qu'un auditeur qualifierait comme étant la meilleure, et donc nous n'avons conservé que ces deux critères dans notre algorithme.

Certaines situations peuvent toujours rester ambiguës, par exemple lorsque deux pulsations de périodes 500 et 1000 ms sont présentes sur l'ensemble de la séquence. Sur une partition, ces pulsations pourraient par exemple être notées par la noire et la croche. Nous considérons alors comme la meilleure la pulsation dont le nombre d'harmoniques est le plus grand. Comme les harmoniques considérées se limitent aux rapports simples de proportion, la pulsation qui potentiellement en possède le plus peut être vue comme un centre de gravité, possédant à la fois des harmoniques inférieures et supérieures. Les différentes pulsations en relation harmonique pourraient par ailleurs être interprétées en termes de différents niveaux métriques. Nous aborderons ce problème de façon plus générale au chapitre suivant.

3.3 Conclusion : principaux avantages par rapport à la méthode de Dixon

Les principaux avantages de notre méthode sont :

- Complexité linéaire : notre algorithme est de complexité linéaire alors que celui proposé par Dixon [15] était exponentiel. Pour cela, nous avons dû traiter l’ambiguïté de façon locale en prenant un choix définitif sur chaque battue de pulsation après avoir dépassé une fenêtre de taille finie alors que Dixon conserve toutes les possibilités en créant une nouvelle pulsation pour chaque ambiguïté rencontrée. Nous affirmons donc que les choix de battues peuvent se faire localement, comme le ferait un auditeur, et non globalement une fois la séquence entièrement parcourue. Par ailleurs, notre modèle, par son traitement local de l’ambiguïté, peut remettre en cause une battue correspondant exactement à la position attendue lorsque l’un des événements suivants, toujours situé dans la fenêtre de tolérance, se trouve mieux correspondre. Au contraire, Dixon ne peut revenir sur le choix d’un événement correspondant exactement avec la battue attendue.
- Temps réel : notre modèle peut être implémenté en temps réel, ce qui va dans le sens d’un modèle s’approchant de nos propres mécanismes de perception de la pulsation.
- Robustesse : nous pouvons induire de nouvelles pulsations à chaque déplacement de fenêtre ce qui permet de suivre des ruptures brutales de tempo, ou de faire l’analyse de morceaux dont le début ne contient pas forcément la pulsation qui sera perçue par la suite, ce qui est le cas par exemple des introductions libres. Par ailleurs, les marquages sont utilisés eux aussi à chaque étape et interviennent à la fois comme support pour l’extraction de la meilleure pulsation, mais aussi comme indices pour l’induction de nouvelles. Enfin, nous prenons en compte les différentes relations proportionnelles existant entre les pulsations ce qui permet par exemple de mieux évaluer leur validité pour l’extraction de la meilleure pulsation.
- Modularité : l’utilisation des marquages rend le système facilement paramétrable et adaptable à différents cas de figure. Par exemple, le marquage1 peut être utilisé pour repérer les notes ‘style batterie’ c’est-

à-dire répétées à intervalles réguliers, ce qui était rapporté comme un manque à l'algorithme de Dixon.

Des exemples généraux d'analyse rythmique seront montrés à la fin du chapitre suivant ainsi qu'au chapitre 7. Nous nous proposons maintenant d'étudier au chapitre suivant un modèle d'extraction de la métrique à partir des informations sur la pulsation dont nous disposons.

Chapitre 4

L'analyse rythmique : métrique et quantification

Ce chapitre traite le problème de l'analyse de la métrique à partir de la connaissance d'une pulsation. Il propose aussi une courte synthèse des différentes méthodes de quantification actuellement disponibles.

4.1 La mesure

Après avoir donné une définition de la mesure, nous partirons des différents algorithmes proposés dans la littérature pour élaborer notre propre système d'extraction automatique de la métrique.

4.1.1 Définitions

D'après la définition de Cooper et Meyer [11] :

« Meter is the number of pulses between the more or less regularly recurring accents »

Nous adopterons cette définition par la suite, ce qui supposera à la fois la connaissance d'une pulsation, acquise grâce à l'algorithme décrit au chapitre précédent, et la connaissance d'une structure d'accentuation, que nous po-



FIG. 4.1: Une mesure à 13 battues de pulsation, l'unité de référence étant la croche (représentée par le chiffre 8)

serons en première étape de notre algorithme d'extraction de mesure. Nous allons maintenant décrire plus précisément ce que représente pour nous la mesure.

La représentation de la mesure dans la notation rythmique traditionnelle

Dans la notation rythmique traditionnelle, des barres de mesure indiquent les groupements de pulsation. Les barres sont accompagnées de deux chiffres, l'un indiquant le nombre de battues de pulsation contenues dans le groupement, l'autre indiquant l'unité de référence de la pulsation. Par exemple, la figure 4.1 montre une notation de mesure à 13 croches ($13/8$).

La pulsation de référence peut différer suivant les groupements, mais elle sera alors le plus souvent multiple des autres pulsations.

Cette notion de groupement est sans doute à relier encore une fois à notre perception du temps. Comme nous ne pouvons compter jusqu'à l'infini, il nous faut regrouper pour trouver une nouvelle unité à partir de laquelle nous pourrions à nouveau compter. Par ailleurs, notre mémoire finie présente l'avantage de forcer les groupements de petites tailles, de l'ordre de quelques pulsations, ce qui augmente les chances de trouver des groupes de tailles similaires ou identiques, donc comparables, et par là favorise l'établissement de relations entre ces groupes. Cet aspect sera par contre assez difficile à modéliser précisément.

La mesure, une possible irrégularité

Contrairement à la pulsation qui se définit par rapport au concept de régularité, le découpage en mesures n'est pas forcément régulier. Cette possible irrégularité peut concerner le nombre de groupements ou bien l'unité du groupement. Il est très difficile d'imaginer une suite de pulsations différentes dans le temps (hormis les variations locales, continues, dont nous avons parlé au chapitre précédent), mais on peut facilement imaginer une suite de mesures de tailles variables.

Par exemple, la musique de Stravinsky utilise ce genre de métrique. Comme le remarque Jacques Rivette [43] :

« Le principe de la musique de Stravinsky, c'est la rupture perpétuelle de la mesure. La grande nouveauté du *Sacre du Printemps* était d'être la première œuvre musicale où le rythme variait systématiquement. A l'intérieur du domaine rythmique, pas de domaine tonal, c'était presque une musique sérielle, faite de l'opposition de rythmes, de structures et de séries de rythmes. »

Certains groupes de jazz comme Aka-moon utilisent les variations métriques de manière systématique, par exemple sous forme d'une progression continue et cyclique du nombre de pulsations par mesure. Dans tous les cas, ces variations seront très difficiles à modéliser, aussi dans un premier temps nous considérerons que la séquence que nous analysons ne contient pas de groupements de mesure irréguliers, même si notre algorithme peut en tenir compte du fait du parcours chronologique de la séquence.

La, ou les mesures ?

Il n'existe pas toujours un seul découpage en mesures d'une même séquence musicale. Suivant les objectifs, le nombre de découpages possibles est potentiellement infini. Par exemple, l'utilisation des mesures peut avoir pour but de simplifier la notation, simplifier l'interprétation, souligner certains accents musicaux. Parfois, les mesures peuvent même changer au fil des

éditions successives de l'œuvre, comme ça a été le cas par exemple pour la 'Danse sacrée', dernier tableau du *Sacre du printemps* de Igor Stravinsky.

Il nous faut alors choisir quel point de vue adopter dans notre recherche. Dans le même esprit que le chapitre précédent, nous nous intéresserons au découpage des mesures tel qu'il serait perçu par un auditeur. Comme la pulsation est supposée donnée, nous supposerons aussi qu'il existe des groupements métriques supérieurs perceptibles.

Les différents niveaux hiérarchiques

Par rapport à la pulsation, la mesure est un concept récursif puisqu'il est possible de grouper des mesures et ainsi de suite de définir une hiérarchie de groupements. La pulsation devient alors l'unité de base à partir de laquelle s'effectuent les groupements successifs.

Généralement, les barres de mesure d'une partition ne matérialisent qu'un seul niveau de groupement au-dessus de la pulsation. Les groupements supérieurs ou inférieurs ne sont que rarement indiqués, pourtant ils sont très souvent perceptibles en musique classique.

La notion de mètre

Un groupement ne se caractérise pas seulement par ses frontières, mais aussi par les différents schémas d'accentuation qu'il contient. C'est ce que nous appellerons le mètre.

Le mot mètre provient du terme grec 'metron', ou alors du terme latin 'mensura' ou 'mensio', qui signifient tous trois 'mesure'. Les définitions peuvent alors provoquer un certain flou dans les concepts qu'elles véhiculent.

De notre côté, nous définirons le mètre comme étant une séquence d'accents répétée au moins une fois. C'est donc un groupement comme la mesure, mais qui a la particularité de contenir une séquence précise d'accents dits 'métriques'.

Les accents métriques

Les accents métriques se distinguent souvent des autres accents tels que les accents mélodiques, harmoniques ou même rythmiques par le fait qu'ils se produisent avec régularité. Le problème est que les différents types d'accents interagissent, donc peuvent 'résonner' ou 's'opposer', et par là il devient difficile de savoir quel type influence l'autre, et plus particulièrement lequel est à l'origine de l'autre : les accents métriques seraient induits à partir des autres accents, mais en même temps ils modifient la perception qu'on a d'eux.

Nous allons maintenant étudier deux approches qui nous semblent très intéressantes pour répondre au problème que nous venons d'introduire.

4.1.2 Les différentes approches

Si la recherche de pulsation est un sujet assez abordé, celle de la métrique l'est beaucoup moins.

Brown [5] propose une méthode par l'autocorrélation pour extraire le niveau métrique correspondant à la notation des partitions. Seule la séquence d'inter-onsets d'un fichier MIDI est considérée (les inter-onsets sont les intervalles de temps entre deux onsets consécutifs et les onsets sont les positions temporelles, souvent exprimées en millisecondes, des différents événements (notes ou accords) du fichier MIDI). Brown part de l'hypothèse que les événements MIDI sont positionnés très fréquemment sur les positions de début de mesure, et comme la mesure peut être vue comme étant un écart temporel régulièrement répété, elle propose de le trouver en appliquant le coefficient d'autocorrélation à la suite d'inter-onsets. L'autocorrélation est une mesure mathématique qui permet de détecter les répétitions contenues dans n'importe quelle séquence de valeurs. Les répétitions se font pour des séquences comportant exactement le même nombre d'événements. Le nombre d'inter-onsets entre deux mesures pouvant changer, l'autocorrélation ne pourrait être directement appliquée aux séquences d'inter-onsets. Brown propose alors une sorte de pré-quantification de la séquence en exprimant chaque valeur d'inter-

onset en fonction d'un nombre plus petit pris comme référence. Le nombre d'éléments des séquences comparées par l'autocorrélation devient alors représentatif des durées temporelles de ces séquences. L'étape de pré-quantification est toutefois difficile à réaliser puisque les divisions des inter-onsets ne sont jamais parfaites ce qui amène à une représentation faussée de la séquence initiale. En fait, on se rend compte qu'il serait nécessaire d'extraire préalablement une pulsation minimale qui serait alors prise comme référence pour le découpage des inter-onsets. Seulement, certains inter-onsets plus petits que la pulsation ne pourraient toujours pas être exprimés. Nous proposerons dans notre méthode de considérer les seuls événements situés sur une pulsation minimale préalablement extraite. Un autre inconvénient de cette approche est que les positions temporelles des mesures ne sont pas détectées. En effet, seules les périodes des mesures peuvent être détectées par l'autocorrélation. Par ailleurs, d'autres paramètres comme les hauteurs et durées ne sont pas considérés dans le calcul.

Cambouropoulos [6] propose de rechercher le mètre en deux temps : tout d'abord en déterminant la structure d'accentuation de la séquence musicale, et ensuite en déterminant la grille métrique la plus appropriée à la structure d'accentuation. La structure d'accentuation est déterminée par des principes de proximité et de similarité issus de la théorie de la gestalt. Un avantage de cette méthode est que contrairement à Brown, plusieurs autres paramètres que les onsets sont pris en compte. Cependant, l'algorithme qui détermine la meilleure grille métrique ne semble pas très approprié à certains cas de figure. Par ailleurs, les variations entre les intensités d'accents ne sont pas prises en compte par sa méthode, alors qu'elles devraient aider à la détermination du mètre par leur succession de temps forts ou faibles.

4.1.3 Conjuguer l'analyse des similarités avec celle des accents

Les deux approches que nous avons présenté ont chacune leurs inconvénients. L'approche de Brown ne prend pas en compte les paramètres musi-

caux autres que les onsets, elle ne détecte pas les positions temporelles des mesures, et elle opère une quantification qui fausse les données. Par ailleurs, Cambouropoulos se base uniquement sur une considération hors contexte des accents pour extraire le mètre alors que la répétition de leurs profils pourrait être détectée avec profit. Nous pensons cependant qu'il est possible de tirer parti des avantages de chaque méthode sans leurs inconvénients en intégrant les deux approches en une seule et même méthode. En effet, chacune pallie à certains inconvénients de l'autre.

L'approche que nous proposons peut donc être vue comme une compilation des approches de Brown et Cambouropoulos, mais si les concepts utilisés sont les mêmes, leur application au matériau musical reste différente.

Nous pouvons distinguer dans notre approche deux étapes : tout d'abord, nous déterminons la structure d'accentuation des événements d'une séquence MIDI. Seuls les événements situés sur des battues d'une pulsation préalablement déterminée sont considérés. La pulsation peut être extraite par exemple à l'aide de l'algorithme proposé au chapitre précédent. Dans un deuxième temps nous extrayons le mètre à partir du calcul du coefficient d'autocorrélation appliqué à cette structure d'accentuation. Ce calcul est possible car les événements considérés sont régulièrement espacés dans le temps. Nous pouvons alors analyser par autocorrélation la séquence d'accentuation qui leur est associée, c'est-à-dire étudier les répétitions des profils d'accentuation. Le fait de considérer les accents permettra d'intégrer à la mesure d'autres paramètres que les inter-onsets.

4.1.4 Première étape, rechercher une structure d'accentuation

L'objectif de cette sous-partie est de déterminer, à partir d'une séquence MIDI et d'une pulsation de référence, l'accentuation métrique de chaque battue de la pulsation.

Contrairement aux marquages employés au chapitre précédent qui s'appliquaient aux événements du fichier MIDI, il s'agit de donner un poids aux

différentes battues de la pulsation selon qu'elles influencent notre perception des accents métriques.

Plusieurs marquages peuvent être choisis. Par exemple, suivant le modèle de Cambouropoulos, nous pourrions marquer les notes à partir de critères gestaltistes appliqués à différentes propriétés MIDI telles que les hauteurs, durées ou intensités.

Nous avons choisi dans notre approche de ne pas considérer les relations qu'entretiennent les différentes battues de pulsations entre elles. Chaque battue a alors été marquée d'après ses propriétés propres, indépendantes de sa place dans la séquence MIDI. Par ailleurs, seuls les événements positionnés exactement sur la battue, et pas ceux compris entre deux battues, ont été considérés. Tous ces choix sont faits pour réduire l'information dont nous disposons de manière à traiter le plus de cas de figure possible. Nous avons alors considéré cinq marquages différents, le principe étant d'accorder un poids fort aux événements perceptivement forts.

- Le marquage M1 pondère une battue proportionnellement à sa dynamique.
- Le marquage M2 pondère une battue proportionnellement à son ambitus, c'est à dire l'intervalle de hauteur entre ses deux hauteurs extrêmes.
- Le marquage M3 pondère une battue qui est suivie d'un silence.
- Le marquage M4 pondère une battue proportionnellement à sa durée.
- Le marquage M5 pondère une battue proportionnellement au nombre de notes qu'elle contient.

Pour chacun des cinq marquages, mis à part M3, un poids normé entre 0 et 8 est donné proportionnellement aux différentes valeurs des propriétés correspondantes. Le marquage M3 qui est booléen est pondéré par 0 ou 8.

Ensuite, les poids des différents marquages sont linéairement combinés battue par battue. Au final, on dispose d'une séquence de poids révélant notre modélisation de la structure d'accentuation. Un exemple d'application des marquages est montré figure 4.3.

4.1.5 Deuxième étape, rechercher un groupement

Notre objectif est maintenant de déterminer, à partir d'une séquence accentuée de battues, les positions des barres de mesures telles qu'elles seraient perçues à l'audition de la séquence.

Comme nous l'avons déjà précisé, nous allons uniquement rechercher des barres de mesure régulièrement espacées. Cela revient alors à déterminer une taille de groupement, que nous exprimerons en nombre de pulsations, couplée à la position du premier groupement.

Cambouropoulos [6] propose de déterminer la meilleure grille métrique que l'on peut appliquer à la structure d'accentuation de la séquence de pulsations. Une grille métrique est définie par une série de positions temporelles régulièrement espacées, caractérisée par son espacement et sa première position temporelle. Chaque position possible de chaque grille envisagée par sa méthode est évaluée par une valeur obtenue en faisant la somme des poids de tous les événements de la séquence placés sur les barreaux de la grille. La meilleure grille est celle dont les différentes positions correspondent aux plus grands écarts de valeur.

Cette approche présente toutefois plusieurs inconvénients : le critère qui détermine les 'plus grands écarts de valeur' n'est pas très bien défini, et dépend de l'appréciation de l'utilisateur. De plus, la meilleure position de grille est calculée à partir des valeurs absolues des poids d'accent sans tenir compte des variations de poids entre les accents. On peut alors se demander si un profil d'accent tel que (0 0 0 1 1 1 2 2 2 1 1 1) pourrait être interprété comme contenant des groupements par trois. En effet, le calcul des valeurs de position pour une grille ternaire aurait pour résultat la valeur 4 ($0+1+2+1$) pour les trois positions possibles. La différence entre les trois valeurs étant nulle, la grille ternaire serait alors considérée comme très médiocre par rapport aux autres grilles possibles. Pourtant, les éléments de cette séquence d'accents pourraient indiscutablement être groupés par trois. Par ailleurs, la méthode proposée ne moyenne pas les différentes valeurs de position par le nombre de barreaux de chaque grille ce qui fait que plus une grille a de barreaux, plus

les différences entre ses différentes valeurs de position ont des chances d'être importantes.

Au final, on peut se demander si les groupements peuvent être déterminés par une mesure statistique globale effectuée sur les poids d'accents. Nous pensons que la succession des poids d'accents dans le temps a aussi son importance et doit être prise en compte par la mesure. Par exemple, une des caractéristique des groupements que nous recherchons est leur périodicité. La répétition régulière de certains profils d'accents dans le temps pourrait alors nous permettre d'évaluer ces périodicités.

Pour cela, nous avons choisi d'utiliser une fonction d'autocorrélation.

4.1.6 La recherche d'un groupement : analyser les répétitions par autocorrélation

La fonction d'autocorrélation a déjà été utilisée par Brown [5], et Desain [12] pour analyser la métrique. Cette fonction permet de repérer des répétitions non exactes à l'intérieur d'une suite de nombres. La méthode proposée par Brown consistait à appliquer directement la fonction sur une séquence d'inter-onsets (un inter-onset est un intervalle de temps entre deux onsets successifs) pour extraire différents niveaux métriques. Cela soulevait différents problèmes puisque la pulsation n'était pas supposée connue, et faisait partie des éléments à découvrir. La séquence pouvait donc être de tempo variable, ce qui était difficilement analysable par l'autocorrélation. En effet, l'autocorrélation ne fait de comparaisons qu'entre des séquences de même taille. Si l'on exprime la taille de la séquence à analyser proportionnellement au temps (n'oublions pas que nous analysons la structure rythmique), tout changement de tempo entraînera un changement de taille de séquence. Prenons par exemple la séquence de 5 noires à la seconde. Supposons que la séquence soit exprimée dans l'unité 100 millisecondes, ce qui pourrait être représenté par (1 0 1 0 1 0 1 0 1 0). Une diminution de tempo de 5/2 noires à la seconde provoquerait un décalage de 5 unités par seconde. La séquence deviendrait alors (1 0 0 1 0 0 1 0 0 1 0 0 1 0 0). Comparée à la séquence

initiale par autocorrélation aucune similarité ne serait détectée. Par ailleurs, nous avons vu que d'autres paramètres que les onsets participent à notre perception de la pulsation et il est nécessaire pour les prendre en compte de réaliser par exemple un marquage de la séquence avant de lui appliquer la fonction d'autocorrélation. Enfin, les phases d'une pulsation ne peuvent pas être détectées par l'autocorrélation qui n'est sensible qu'aux périodes.

Toutes ces remarques ne sont plus valables dans notre contexte puisque nous connaissons une pulsation. La séquence est donc facilement régularisable par compression-dilatation rythmique de chaque intervalle de temps entre chaque battue de pulsation afin d'éliminer les changements de tempo. Par contre, la fonction d'autocorrélation n'étant pas sensible aux phases, nous utiliserons une autre méthode pour déterminer la position de début du premier groupement. Nous allons maintenant montrer comment extraire à l'aide de la fonction d'autocorrélation la meilleure taille de groupement d'une séquence de poids d'accents.

Définition

Etant donnée une séquence $x[n]$ de M valeurs (on considère que M est aussi grand que nécessaire), et un entier m tel que $0 \leq m \leq M$, la valeur d'autocorrélation $A(m)$ entre la séquence $x[0..N]$ et la séquence $x[m..m+N]$ est donnée par la formule :

$$A(m) = \frac{\sum_{n=0}^{N-1} (x(n) \times x(n+m))}{\sum_{n=0}^{N-1} x(n)^2} \quad (4.1)$$

avec $N = M-m$

Plus la valeur d'autocorrélation est grande, plus les séquences $x[0..N]$ et $x[m..m+N]$ sont similaires. A partir des $N+1$ valeurs d'autocorrélation calculées sur la séquence $x[0..N]$, on sélectionne celles qui sont maxima locaux dans une fenêtre de taille donnée (voir l'exemple donné figure 4.2). Cela revient à repérer les sous-séquences qui sont le plus corrélées avec $x[0..N]$. Pour définir si une valeur située dans la séquence à la position p est maximum local, on

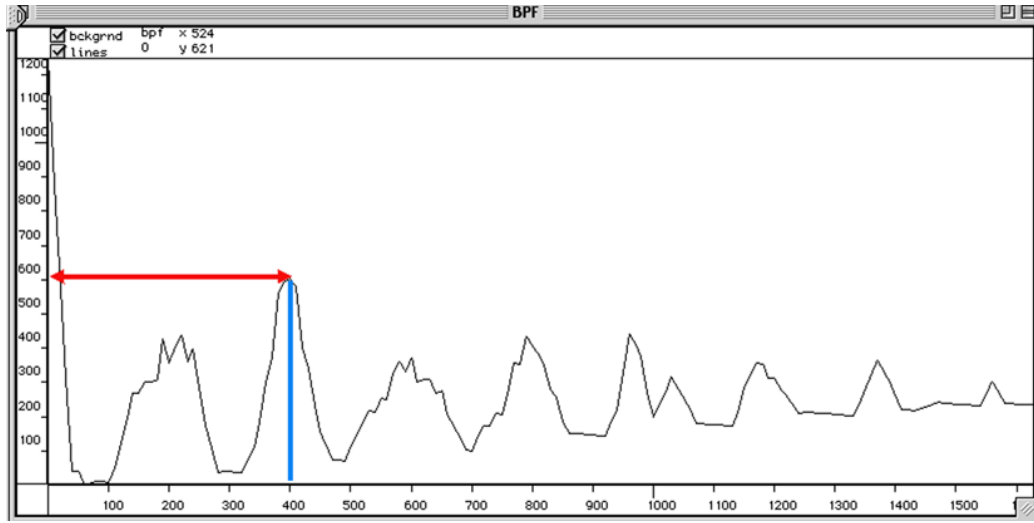


FIG. 4.2: Un graphe d'autocorrélation. Les pics localement maximaux sont interprétés comme débuts des sous-séquences les plus auto-corrélées avec la séquence analysée.

place une fenêtre centrée sur p que l'on divise en deux zones : une petite zone centrée sur p de taille $1/3$ de p , et une grande zone centrée sur p de taille p . La valeur à la position p est maximum local si la valeur correspondante d'autocorrélation est maximale dans la petite zone, et au moins supérieure à $1/3$ de la plus grande des valeurs située dans la grande zone.

Le résultat du calcul d'autocorrélation fournit plusieurs valeurs de maxima locaux qui indiquent la forte corrélation entre la séquence $x[0..N]$ et les séquences $x[m..m+N]$, $x[n..n+N]$. Les positions m , n des séquences corrélées sont alors interprétées comme périodes possibles de la séquence globale $x[n]$ sur laquelle on a fait l'analyse.

Application du coefficient à la recherche d'une taille de groupement

En supposant que l'on applique directement la formule d'autocorrélation à notre séquence de poids Marqu $[0..N]$, nous obtiendrions les positions $(m, n, \dots)$ des sous-séquences les plus corrélées. Cependant, la formule ne prend pas en compte les périodicités possibles des séquences telles que Marqu $[1..1+N]$

... Marqu[k..k+N] qui ne commencent pas au début de Marqu[0..N], mais à partir des rangs 1..k. En effet, il est possible qu'un groupement ne commence que quelques positions après le début de la séquence de poids. Nous devons donc appliquer la formule d'autocorrélation à partir de tous les débuts possibles, soit N+1 fois (une application similaire de la formule a été proposée par Desain [12] pour mesurer l'expressivité musicale). Au final, nous obtenons une liste de périodes pour chaque début possible. Ce parcours chronologique de la séquence permettrait en plus d'analyser les variations de groupements métriques dans le temps, ce qui concerne un certain nombre d'œuvres musicales. Cependant, dans un premier temps, nous n'avons considéré que le point de vue global, c'est à dire une seule métrique (une seule période) pour toute la séquence.

Pour déterminer alors quelle est la période la meilleure, nous classons toutes les périodes pour tous les débuts possibles suivant leur nombre d'occurrences. La période 'la plus fréquente' de la liste est finalement considérée comme étant la meilleure. Un exemple d'analyse du début du deuxième mouvement de la Sonate en Do Majeur de Mozart est donné figure 4.3. Les périodes les plus fréquentes sont les suivantes : ((4 71) (2 49) (24 27) (12 21) (8 14) (3 12) (1 6) (20 5)). Le groupement par 4 est trouvé comme étant le plus plausible (il est proposé pour 71 évènements de la séquence), suivi du groupement par 2 (proposé pour 49 évènements).

Il est ensuite possible d'itérer le processus à nouveau à partir des groupements trouvés, de façon à chercher des groupements de niveaux supérieurs.

En supposant par exemple que les battues peuvent être groupées par deux, les deux étapes de notre algorithme, c'est à dire le marquage et la recherche du mètre, peuvent être appliquées de nouveau sur chaque début des nouveaux groupements. Ainsi de suite nous pouvons appliquer le processus jusqu'à ce que le groupement proposé soit de taille un, ce qui signifiera que l'algorithme ne peut plus prolonger l'analyse.

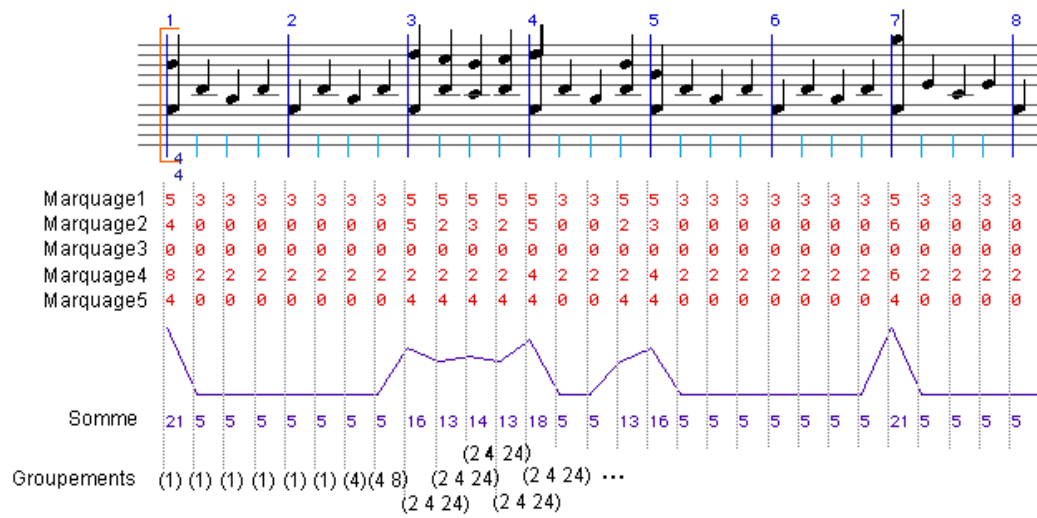


FIG. 4.3: Analyse métrique du début du 2ème mouvement de la sonate en Do majeur de Mozart. Les 5 marquages apparaissent en haut de la figure (une ligne par marquage). En bas apparaissent les différents groupements trouvés pour chaque évènement (groupements par 1, puis groupements par 4 et 8 puis groupements par 2, 4 et 24). Sur l'ensemble de la séquence, le groupement par 4 est celui qui apparaît le plus fréquemment.

4.2 La quantification des durées en intervalles proportionnels

Le problème de la quantification, à savoir trouver une notation rythmique satisfaisante à partir d'une suite d'évènements dont les positions et les durées sont exprimées dans une échelle de temps très (trop) fine, par exemple la milliseconde, peut maintenant être résolu par la donnée d'une pulsation et d'une métrique extraites à l'aide des algorithmes que nous venons de présenter. En effet, le problème se réduit à effectuer la quantification pour chaque segment de pulsation, comprenant au plus une dizaine d'évènements, ce qui ne pose en général pas de problèmes pour les quantificateurs actuels. Ainsi, notre approche propose une réponse au problème de la propagation croissante des erreurs de quantification en passant par l'extraction préalable d'une pulsation et d'une métrique à partir desquelles seulement nous entamons une quantification. Nous allons maintenant présenter brièvement quelques quantificateurs généralement plus performants que les quantificateurs traditionnels. Ces quantificateurs pourront être utilisés pour quantifier séparément chaque segment de pulsation déterminé par notre algorithme.

4.2.1 Définitions

La grille temporelle définie par la donnée de la pulsation et la mesure offre une structure, un squelette par rapport auquel nous pouvons représenter les onsets et les durées d'une séquence musicale qui vont 'l'enrober'. Pour cela, il est nécessaire d'opérer une quantification, c'est à dire qu'il faut modifier les onsets et les durées de manière à ce qu'ils soient en proportion des positions de la grille temporelle.

Lorsqu'il est nécessaire d'approximer les durées pour faire apparaître des rapports de proportion, on effectue une quantification. Cette étape est couramment effectuée par les logiciels du commerce (éditeurs graphiques, séquenceurs, logiciels de gravure), mais les résultats sont rarement satisfaisants. Certains quantificateurs s'inspirent alors du domaine de l'intelligence

artificielle pour appliquer à la musique ses concepts.

4.2.2 Quelques modèles de quantification

Desain et Honing [13] proposent un quantificateur basé sur un réseau à satisfaction de contraintes sans apprentissage. Chaque cellule du réseau correspond à l'une des durées à quantifier. Des cellules représentant des durées voisines interagissent entre elles par l'intermédiaire d'une fonction d'interaction qui les modifie de manière à ce que les durées aient 'la meilleure proportion possible'. Une fois le réseau stabilisé, les durées quantifiées sont envoyées en sortie du réseau.

Un autre quantificateur proposé par Longuet-Higgins [25] repose sur un modèle symbolique basé sur une connaissance minimale du mètre. Plusieurs pulsations sont testées de la façon suivante : Une pulsation calée sur le début de la séquence à quantifier est subdivisée en deux ou trois de façon à trouver la meilleure correspondance avec les onsets des premières durées. L'algorithme itère ensuite le processus, positionnant séquentiellement les pulsations jusqu'à aboutir à la fin de la séquence. Eventuellement, suivant la subdivision choisie, la pulsation peut être modifiée au cours de l'itération. Le résultat est une suite de pulsations subdivisées en deux ou trois. Plusieurs pulsations sont testées, et la meilleure suite de subdivisions est retenue.

Ces deux algorithmes sont assez représentatifs des méthodes de quantification disponibles car ils illustrent deux différentes approches caractéristiques des problèmes posés en intelligence artificielle.

Pour mieux comprendre leur qualités et défauts, les modèles doivent être décrits sous différents points de vue, ce qui a été fait dans [Desain 93] pour le cas des deux quantificateurs dont nous venons de parler. Ces quantificateurs permettent de mesurer le potentiel que peut apporter le domaine de l'intelligence artificielle à la modélisation rythmique. Le résultat final ne peut cependant toujours être satisfaisant, car il dépend des attentes de l'utilisateur en matière par exemple de précision ou de cohésion dans la notation. Les quantificateurs sont alors souvent paramétrables de manière à prendre

en compte les préférences individuelles.

4.3 Exemples

Les algorithmes présentés au chapitre précédent et dans celui-ci ont été utilisés pour quantifier automatiquement une base de données de 500 fichiers MIDI dont nous allons nous servir au cours des chapitres suivants. Par ailleurs, différents tests spécifiques de validation ont été effectués à partir de corpus différents de manière à mettre à l'épreuve notre logiciel.

Nous avons ainsi testé la robustesse de notre modèle aux changements de tempo en analysant quatre Ragtimes auxquels nous avons appliqué 5 déformations temporelles d'intensité croissante ¹. Les résultats sont montrés figure 4.4. En haut sont représentés les graphes des variations de tempo qui ont été appliquées aux morceaux dont le tempo était initialement constant. Pour chaque déformation et pour chaque morceau correspond alors un graphe représentant l'évolution des périodes de la l.p.p dans le temps. Des profils similaires aux déformations appliquées montrent que les variations ont bien été détectées par l'algorithme. Le seuil à partir duquel les résultats de l'algorithme ne sont plus acceptables se situe au niveau de la dernière déformation (la plus importante). Ce genre de déformation, de par son ampleur, se rencontre rarement dans la réalité. La figure montre par ailleurs que les périodes petites résistent mieux aux déformations que les grandes, ce qui s'explique par le fait qu'elles réagissent plus vite du fait de la fréquence plus grande de leur nombre de battues.

Nous avons par ailleurs comparé notre algorithme avec les résultats rapportés par Dixon [17] et Cemgil [7] sur un ensemble de 219 interprétations différentes de 'Michelle' et 'Yesterday' des Beatles. Nos résultats sont présentés figure 4.5, à comparer avec ceux de Cemgil (probabilistic) et Dixon (multi-agent) figure 4.6. Nos temps de calcul sont satisfaisants, de l'ordre

¹les exemples sonores sont disponibles à l'adresse
<http://www.ircam.fr/equipes/repmus/meudic>

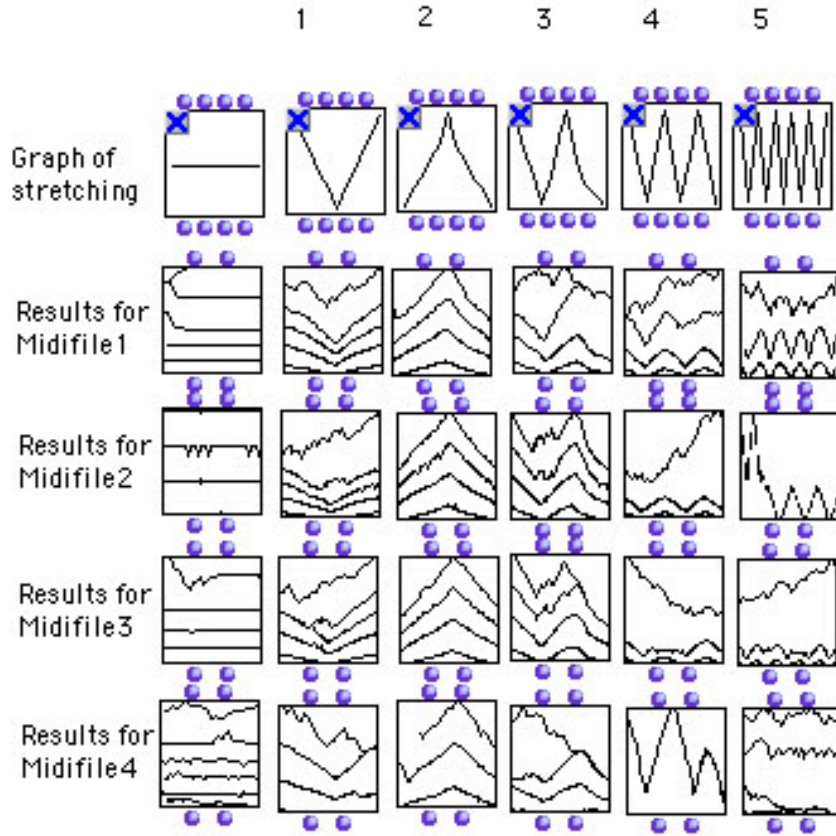


FIG. 4.4: Quatre ragtimes auxquels nous avons appliqué cinq déformations de tempo d'intensité croissante (de gauche à droite). La l.p.p issue de l'algorithme est représentée pour chaque cas de figure. Les résultats sont satisfaisants, et montrent les limites de l'algorithme lorsque les changements de tempo deviennent trop importants (graphes de droite).

	Yesterday	Michelle
<i>By subject group</i>		
Professional jazz	97 +/- 1	97 +/- 1.5
Amateur classical	95 +/- 2.5	96 +/- 2
Professional classical	92 +/- 3	88 +/- 5
<i>By tempo condition</i>		
Fast	95 +/- 2	93 +/- 5
Normal	94 +/- 3	94 +/- 4
Slow	94 +/- 3	93 +/- 5
Average	94 +/- 3	93 +/- 5

FIG. 4.5: Table de résultat des tests effectués sur 219 interprétations de Michelle et Yesterday des Beatles. Les tables sont à comparer avec celles rapportées par deux autres méthodes figure 4.6.

de 1/2 seconde par morceau (tous les morceaux sont de même taille). Dixon rapportait un temps entre 2 pour cent et 10 pour cent de la taille du morceau, ce qui correspond pour les Beatles à une fourchette de 1 à 5 secondes. Les différents résultats sont très similaires à ceux rapportés par Dixon et Cemgil. La catégorie ‘pianiste professionnel’ a de moins bons résultats que les autres, ce qui était aussi rapporté par les deux autres approches.

Enfin, figure 4.7 nous montrons le résultat d’une analyse de la métrique à partir du morceau ‘Take5’ de Paul Desmond. Le groupement par cinq est proposé par l’algorithme ce qui ne contredit pas le titre du morceau. De nombreux autres exemples concernant l’extraction de la métrique peuvent être trouvés dans Meudic [33].

	Yesterday Data Set		Michelle Data Set
	Probabilistic	Multi-agent	Multi-agent
<i>By subject group</i>			
Professional jazz	95 ± 3	95 ± 7	95 ± 2
Amateur classical	92 ± 8	93 ± 13	94 ± 4
Professional classical	89 ± 7	95 ± 4	88 ± 12
<i>By tempo condition</i>			
Fast	94 ± 5	96 ± 3	94 ± 8
Normal	92 ± 8	95 ± 9	93 ± 6
Slow	90 ± 7	92 ± 12	91 ± 9
Average	91 ± 7	94 ± 9	93 ± 8

FIG. 4.6: Tables rapportées par Cemgil(probabilistic) et Dixon (multi-agent) à comparer avec les notre figure 4.5.

Measure :	Position of first measure :
5/4	4
10/4	4
3/4	1

FIG. 4.7: Proposition de métrique pour 'Take5' de Paul Desmond.

4.4 Récapitulatif

Ce chapitre et le précédent ont été consacrés à l'analyse rythmique, et nous avons proposé plus particulièrement deux algorithmes complémentaires, l'un d'extraction automatique de la pulsation, et l'autre d'extraction automatique de la métrique. Les deux algorithmes produisent un résultat après un temps de calcul assez faible, de l'ordre de $1/25$ du temps total de la séquence à analyser, la plus grande partie du temps étant consacrée à l'extraction des pulsations, 8 fois plus lente que l'extraction de la métrique (pour un processeur G4 466Mhz). Les résultats sont presque toujours significatifs. Lorsque des erreurs se produisent, elles correspondent souvent à des passages dont les temps sont faiblement marqués. Ce type de rythme pourrait être mieux analysé si des marquages spécifiques étaient utilisés, permettant par exemple de rendre compte du swing ou des accents contredisant la pulsation. De fait, alors que presque tous les morceaux de musique classique sont correctement analysés, les morceaux de jazz sont moins bien traités. Une improvisation de Thelonious Monk, par exemple, produira une courbe de tempo assez instable, oscillant entre les notes accentuées situées près des pulsations et les notes correspondant aux pulsations.

La majorité des résultats restant très satisfaisante, nous nous proposons alors de nous baser sur notre analyse rythmique automatique de 500 fichiers MIDI allant de la période baroque (Rameau) au début du 20ème siècle (Ravel) pour entamer une analyse motivique qui sera l'objet des deux chapitres suivants.

Chapitre 5

L'analyse motivique : la mesure des similarités

Ce chapitre est consacré à l'analyse motivique automatique, c'est à dire à l'extraction automatique de motifs à partir d'une séquence musicale donnée. Ce domaine représente une étape importante dans l'analyse musicale puisque d'une part le langage motivique est présent dans un très grand nombre d'œuvres musicales et d'autre part les motifs déterminent très souvent la structure de ces œuvres.

Plus particulièrement, nous nous intéressons dans ce chapitre au problème de trouver une représentation des données d'analyse permettant de réduire la complexité combinatoire inhérente à l'analyse motivique. Pour cela, nous introduisons la notion d'épure, et nous opérons un filtrage de certains éléments de façon à réduire la taille des séquences à analyser. Nous définissons ensuite une mesure de similarité que nous pourrons appliquer à n'importe quelle séquence polyphonique ou monophonique.

Dans le chapitre suivant, nous utiliserons nos résultats pour élaborer une réponse au problème de l'extraction de motifs à partir d'une matrice de valeurs de similarité. Dans une deuxième partie, nous redéfinirons notre mesure de similarité pour répondre au problème complexe mais combinatoirement plus simple de la recherche d'un motif donné, polyphonique ou non, à l'inté-

rieur d'un corpus de plusieurs fichiers musicaux.

Dans ce chapitre après avoir donné notre définition de ce qu'est un motif, nous présenterons l'état de l'art dans le domaine, puis nous construirons notre fonction de similarité.

5.1 Qu'est-ce qu'un motif?

Par rapport au rythme que nous avons abordé dans le chapitre précédent, le motif pourrait constituer l'un des autres composants essentiels à la musique, mais cela reste délicat à établir puisque les motifs ne sont presque jamais explicités par le compositeur sur la partition. A priori, n'importe quelle séquence de notes de taille inférieure à celle de la partition entière est alors susceptible d'être un motif. Cette absence d'indication donne au motif un aspect un peu mystérieux, comme s'il s'agissait d'un secret bien gardé que l'écoute et l'analyse auraient pour rôle de découvrir. Peut-être aussi cette part de mystère et d'ambiguïté participent-elles au plaisir inhérent à l'appréhension de l'œuvre musicale. Une conséquence pour notre recherche est qu'il n'existe pas de motif a priori ce qui nous amènera à déterminer nos propres critères pour juger si tel motif est pertinent ou non pour l'œuvre analysée.

5.1.1 La répétition comme invariant

Alors qu'un rythme ne peut se concevoir sans la notion du temps, un motif se conçoit difficilement sans la notion de répétition. On pourrait imaginer le cas d'un motif isolé dans la partition, jamais répété, par exemple caractérisé par un rythme contrastant, une position entourée de silences, ou un timbre unique. Mais cet élément nous apparaîtra alors différent, détaché de tout le reste, et il incarnera la négation extrême de la notion de répétition, donc restera lié à cette notion. On pourrait dire que les séquences pouvant être à l'origine de motifs sont soit celles qui sont répétées soit celles qui ne sont jamais répétées. Dans un premier temps nous nous focaliserons seulement sur les séquences répétées, car elles restent selon nous à l'origine de la majorité



FIG. 5.1: Deux groupes de notes très similaires (deux accords mineurs transposés de un ton), mais rarement jugés comme tels.

des motifs d'une partition.

Il nous restera à définir ce que l'on entend par répétition motivique, ce qui représente l'essentiel du problème de l'analyse motivique. En particulier, il nous faudra déterminer quelles répétitions dans la partition peuvent être associées à la notion de motif, et dans le cas où différentes répétitions seraient envisageables, nous demander quel point de vue adopter pour en sélectionner seulement une partie que nous jugerons intéressante. Par exemple, la figure 5.1 montre une partition sur laquelle deux groupes de notes ont été sélectionnés : (mi sol si) et (la ré fa). Ces deux groupes peuvent être assimilés à deux accords mineurs, l'un étant dérivé de l'autre par transposition de un ton. Ces deux groupes sont donc similaires, ils constituent une répétition. Pourtant, peu de personnes jugeront ce point de vue satisfaisant, et il nous faudra alors définir précisément ce qu'est pour nous une répétition intéressante.

5.1.2 Le motif, une notion perceptive

Un critère souvent utilisé en analyse pour établir si un motif est la répétition d'un autre est celui de la similarité perceptive. La similarité cognitive

joue aussi sans doute un rôle important, mais comme pour le rythme, il est difficile de faire la part entre notre perception et notre cognition du phénomène de répétition si bien que nous ne parlerons que de perception, en incluant implicitement la cognition dans ce terme.

La perception comme filtre à motifs

Très souvent, les motifs contenus dans une pièce musicale, donc aussi les répétitions, peuvent être perçus par l'auditeur. Ils ont souvent été écrits par le compositeur à cette fin. Une approche pour l'analyse des motifs est alors de détecter les répétitions qui sont perceptibles dans la séquence musicale. En faisant ce choix, toutes les répétitions et similarités possibles contenues dans la séquence musicale ne seront plus également valables car seules celles pouvant être jugées perceptives feront sens pour nous. C'est une façon de limiter le nombre de motifs que nous trouverons, et aussi de rendre les résultats intéressants pour une majorité d'utilisateurs familiers des critères perceptifs que nous emploierons et que nous essaierons de choisir les plus généraux possibles.

Il nous faut donc modéliser les critères perceptifs qui nous permettent généralement d'affirmer que deux séquences données sont proches. Ces critères seront à chercher du côté de la perception propre au phénomène physique musical lui-même, qui se matérialisera pour nous en une séquence MIDI. Ce sera la seule donnée dont nous disposerons pour notre analyse. Cependant, certains aspects externes au phénomène physique pourraient aussi être pris en compte : le vécu et la culture propres à chacun peuvent en effet influencer notre perception. On peut alors se demander dans quelle mesure il nous faut considérer ces aspects.

Dans quelle mesure la culture influe-t-elle notre perception ?

L'évolution de la musique peut être vue comme un phénomène assez continu sans cesse alimenté par notre écoute des musiques du passé et motivé par une envie de les dépasser. La tonalité par exemple est un langage

dont la maturation a été très lente et qui n’a cessé de se développer pendant plusieurs siècles jusqu’à arriver à son seuil de rupture. Cette évolution s’est faite grâce à une culture et une expérience sans cesse enrichies et transmises par le biais d’écrits, de concerts ou par des formes de communication orale.

Faut-il considérer cette part de savoir culturel propre à chacun dans notre recherche des répétitions ? Par exemple, peut-on espérer modéliser un système qui comprenne ce qu’est l’harmonie et qui puisse apprendre le style et les règles propres à chaque époque ?

Nous jugeons cette tâche extrêmement complexe, et nous préférons donc supposer l’existence de règles perceptives simples mais peu spécifiques à une époque particulière pour élaborer notre modèle. Par exemple, nous éviterons de considérer la tonalité comme un acquis, et les notions de cadence, accord parfait ou modulation ne seront par exemple pas utilisées. Nous essaierons à travers nos exemples musicaux de montrer que nos règles voulues ‘a-culturelles’ peuvent aboutir à des résultats musicaux déjà très satisfaisants.

5.1.3 Comment modéliser un motif polyphonique ?

Le terme motif est habituellement appliqué aux séquences monophoniques. Cependant, il est tout aussi valable pour des séquences polyphoniques (voir par exemple l’extrait proposé figure 5.2). Cela s’impose même parfois dans certains cas de figure. Plutôt que de rechercher à l’intérieur de la polyphonie les répétitions des différentes lignes monophoniques, il faut alors rechercher les répétitions entre des fragments entiers de la séquence polyphonique.

Prenons l’exemple de l’extrait figure 5.3 de la neuvième sonate de Beethoven. La séquence constituée par les quatre premiers éléments peut être considérée comme un motif, puisqu’elle est répétée plusieurs fois par la suite. Cette séquence est polyphonique (une note suivie par trois accords de trois notes). Est-ce qu’il faudrait la considérer comme la superposition de trois lignes mélodiques différentes, ou comme une seule entité polyphonique ?

Cela soulève aussi le problème de l’indépendance de la ligne mélodique

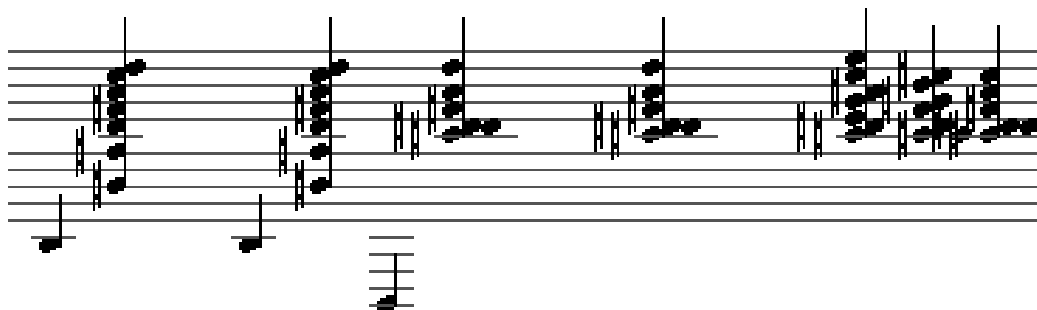


FIG. 5.2: Un extrait de la ‘Danse sacrée’ du *Sacre du printemps* de Igor Stravinsky. Une séquence monodique pourrait difficilement être isolée de la polyphonie.

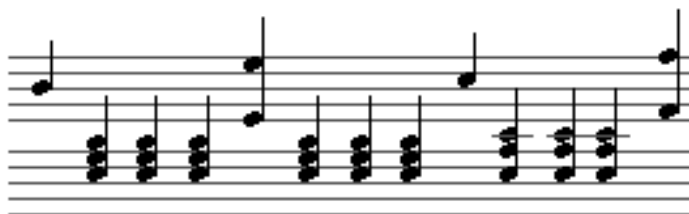


FIG. 5.3: Début de la sonate pour piano N°9 en Mi Majeur de Beethoven. La première note et les trois accords suivants forment un motif. Transformer ces trois accords en trois voix indépendantes ne serait pas très utile pour l’analyse.



FIG. 5.4: Début de l'*Intermezzo* op117 no1 en Mi bémol Majeur de Brahms. La séquence peut être divisée en deux parties. Le contexte polyphonique participe à la compréhension de la structure. La mélodie ne serait sans doute plus perçue si les accords étaient changés.

telle qu'on l'entend habituellement par rapport à son contexte polyphonique.

Souvent, la seule considération de la ligne mélodique ne suffit pas à rendre compte de tous les aspects perceptifs qui seront utilisés par l'auditeur pour établir les similarités. Prenons par exemple l'*Intermezzo* op117 no1 de Brahms figure 5.4. Il serait possible d'extraire à la main la ligne mélodique après avoir entendu le morceau, mais sa seule donnée ne serait pas suffisante pour analyser les répétitions qui sont perçues en partie grâce à la disposition des accords. Une autre disposition des accords, moins régulière et telle que la même ligne mélodique ne serait pas perçue, aboutirait à une analyse différente.

Par ailleurs, la ligne mélodique ne contient pas d'informations, à part le rythme, qui permettraient d'isoler de façon certaine les répétitions. Souvent, on considère pourtant que la mélodie est l'aspect le plus important de l'analyse, mais il s'agit alors le plus souvent de la mélodie et de son accompagnement harmonique implicite. Assez paradoxalement, il nous est difficile de penser une mélodie sans accompagnement harmonique, mais aussi il nous

est impossible de chanter plus d'une note à la fois, si bien que lorsque nous chantons ou pensons une mélodie, nous la sous-entendons avec un accompagnement harmonique même s'il n'est pas explicite. Cet aspect est difficile à modéliser et il fait intervenir des notions comme l'harmonie que nous ne voulons pas utiliser, donc nous n'allons pas le considérer directement. De même, nous n'allons pas chercher à isoler une mélodie, puisque d'une part elle ne pourrait suffire à l'analyse des répétitions, et d'autre part cette tâche pourrait être plus difficile que l'analyse des répétitions elle-même. En effet, si nous reprenons l'exemple de l'Intermezzo de Brahms, nous remarquons que la ligne mélodique n'est pas située aux extrémités, critère souvent évoqué pour repérer automatiquement les lignes mélodiques, mais se situe au milieu de la polyphonie. Ce cas est très fréquent en musique, et certains compositeurs comme Ravel, Fauré ou Wagner l'utilisent presque systématiquement. De plus, elle n'est pas isolée des accords, mais au contraire elle en fait partie. Si perceptivement elle est très facile à isoler, il est difficile de concevoir un algorithme qui pourrait effectuer la même tâche.

Cela dit, les différents éléments de la polyphonie ne jouent pas tous un rôle dans la perception des répétitions. Il serait donc intéressant d'extraire les seuls éléments de la polyphonie intervenant dans notre perception des répétitions. Nous allons donc analyser les répétitions de motifs polyphoniques inclus dans une autre séquence polyphonique.

Pour finir, un motif polyphonique peut être monophonique. Par exemple, les lignes mélodiques des fugues de Bach peuvent être difficiles à isoler, mais toujours elles apparaissent les unes après les autres, ce qui nous permet de les percevoir sans doute plus facilement. A l'exception des fugues polythématiques, le sujet apparaît même souvent seul, puis il est répété et varié dans un contexte polyphonique. Nous serons dans ce cas amenés à comparer une mélodie (le sujet) avec une séquence polyphonique (la fugue), ce qui ne sera toutefois possible que parce que la mélodie est apparue une fois isolée du reste de la séquence. Dans les autres cas de figure, les événements superposés au motif initial devront être pris en compte.

5.2 Différentes algorithmes d'analyse motivique

La littérature compte très peu d'algorithmes (pour ne pas dire aucun) se proposant d'extraire des motifs à partir de séquences polyphoniques.

5.2.1 L'analyse de fichiers audio

Une approche, imaginée pour analyser des fichiers audio, est proposée par Peeters [38]. Le signal est vu comme une succession d'états correspondant à autant d'éléments d'une structure. Plusieurs niveaux hiérarchiques d'états peuvent être considérés. Une matrice de similarité représente des vecteurs multidimensionnels de propriétés calculées échantillon par échantillon. A partir de cette matrice sont détectées les variations des propriétés les plus grandes et les plus rapides, que l'on associe à la notion de frontière entre différents états. Ces états sont ensuite classifiés selon leurs similarités. Cette approche propose une démarche assez originale directement applicable à de la musique polyphonique, mais si elle peut convenir pour des signaux comportant de brusques changements (par exemple de timbre ou d'énergie), elle devrait donner de moins bons résultats pour des musiques dont les répétitions ne sont pas marquées par de brusque changements d'état, comme c'est souvent le cas par exemple en musique classique.

5.2.2 La complexité informative

Levy [23] propose d'évaluer la complexité perceptive de transformation d'un motif en un autre à partir de la théorie de l'information. Il distingue la complexité analytique, mesurable par la quantité d'information contenue dans une chaîne d'évènements, et invariable quels que soit l'ordre des évènements, de la complexité perceptive, variable suivant l'ordonnancement et donc la structure des évènements de la chaîne. Pour cela, il définit un certain nombre de transformations possibles, comme la transposition, symétrie ou rétrogradation qu'il associe à un coût, puis recherche la transformation la moins coûteuse selon différents profils d'auditeurs. Les transformations

s'appliquent à toute une séquence et peuvent être combinées entre elles. Le problème est que les exemples musicaux qui sont proposés semblent réalisés manuellement sans justification ce qui rend le calcul empiriste. Par exemple, certains découpages des incises de la cinquième symphonie de Beethoven sont réalisés de façon ‘immédiate’, en ‘isolant’ des sous-périodes et ‘notant’ les transformations. Un tel découpage, même si trivial pour nous, serait complexe à effectuer par un ordinateur sans qu’aucune indication autre que les transformations possibles ne lui soit donné. Par ailleurs, la méthode s’applique à des monodies et ne propose pas de transformations locales au motif, ni de groupe à groupe, par exemple pour faire correspondre un groupe de deux notes à un groupe de trois notes. Elles restent donc trop simples pour pouvoir être appliquées de manière générale à un extrait musical.

5.2.3 La programmation dynamique

Rolland [44], un peu de la même manière que Levy, propose un modèle de transformation monophonique basé sur la programmation dynamique. À partir d’un fichier MIDI, toutes les paires possibles de motifs sont comparées en recherchant la transformation de coût minimum permettant le passage d’un motif à un autre. Les paires dont la valeur de similarité dépasse un certain seuil sont représentées dans un graphe de similarité. Une optimisation intéressante du temps de calcul est proposée. Cependant, la fonction de similarité utilisée est linéaire, ce qui ne semble pas cohérent avec nos propres critères perceptifs. En effet, cela voudrait dire qu’un motif se constitue par concaténation de plusieurs entités indépendantes sans que les relations entre ces entités ne soient considérées. Nous discuterons de cette notion dans notre section sur les règles perceptives que nous avons adoptées. Enfin, l’algorithme ne peut extraire de motifs à partir d’une musique polyphonique dont les différentes voix ne seraient pas connues.

5.2.4 Une approche issue de la linguistique

Clausen, Engelbrecht, Meyer et Schmitz [8] proposent le système PROMS, un algorithme de recherche exacte de motifs polyphoniques dans une base de fichiers MIDI polyphoniques. Les pulsations et la métrique sont supposées connues. Les seules variations autorisées par le modèle sont la transposition. Certaines notes du motif recherché peuvent devenir floues, c'est à dire qu'elles peuvent correspondre à l'une des notes d'un ensemble donné. L'algorithme est conçu à partir de la méthode 'inverted file index' habituellement employée pour la recherche de mots dans les textes. Cette méthode est appliquée à la musique moyennant quelques adaptations : l'unité de base devient la note, donc le vocabulaire devient fini (127 notes), et l'ordre entre les notes du motif recherché est à prendre en compte dans la recherche. Le problème de cette méthode est que si l'ordre entre les notes est pris en compte, leurs relations intervaliques ou rythmiques ne le sont pas : les notes sont comparées indépendamment les unes des autres. Une notion étendue de la similarité deviendrait alors difficile à modéliser. Par ailleurs, l'approche est destinée à la recherche du motif lui-même et non pas de ses variations. Si le motif recherché peut être différent de celui contenu dans la base de donnée (suite à une question erronée) ses variations possibles (au sens variations intentionnelles du compositeur) ne sont pas pour autant modélisées par l'algorithme. Par ailleurs, le motif est supposé connu à l'avance. Ces types d'algorithmes, dits de 'query by humming', ne peuvent donc directement répondre au problème d'extraction de motifs.

5.2.5 La modélisation des accents

Temperley [49] propose d'appliquer une méthode de pondération pour segmenter un fichier MIDI. Pour cela, il pondère les événements saillants pouvant correspondre à des débuts ou fins de motifs. Par exemple, il repère les grands intervalles d'inter-onsets, ou les longs silences. Il choisit de segmenter des phrases comprenant de préférence huit notes, et contraint les

éventuels motifs à débiter sur des positions métriques similaires, c'est à dire par exemple en début ou en fin de mesure. Cependant, ces règles s'avèrent insuffisantes pour détecter de façon générale une structure motivique, et plus particulièrement les phrases qui constituent déjà une notion très difficile à formaliser. Le problème de la polyphonie s'ajoute alors à ces écueils. Une tentative de le contourner en séparant les différentes voix se montre assez peu convaincante. Au final, on se rend compte que l'un des principaux problèmes de l'analyse motivique, à savoir le parallélisme, n'est pas abordé et se trouve sans doute à l'origine des maigres résultats. Les exemples musicaux utilisés montrent d'ailleurs tous un découpage de la partition en entités similaires sans que ce concept ne soit discuté par la méthode. Par ailleurs, la question de la polyphonie est présentée comme réductible à un problème de séparation des voix, qui se révèle tout autant sinon plus délicat à traiter. Par contre, les règles de pondération pourraient fournir quelques pistes pour préciser les débuts et fins de motifs, mais à condition qu'elles soient associées à une recherche des similarités.

5.2.6 La visualisation dans l'espace

Meredith, Lemström, et Wiggins [29] associent la polyphonie à un problème multidimensionnel. L'algorithme proposé recherche différents appariements des figures géométriques formées par des vecteurs de propriétés en opérant des translations de ces figures dans l'espace multidimensionnel des propriétés. Cette approche est intéressante car elle propose une réponse graphique au problème musical. Cependant, pour le moment seules les répétitions exactes peuvent être détectées. Pour éviter de revoir dans sa globalité le processus de recherche des répétitions, une solution pourrait consister à trouver des représentations abstraites telles qu'à un certain niveau, les séquences similaires pourraient devenir identiques. Le problème de trouver de telles représentations reste cependant épineux.

5.2.7 La modélisation du processus d'induction

Une méthode proposée par Lartillot [21] prend comme objet d'étude un fichier MIDI à partir desquels sont induits des motifs dans un ordre chronologique. Toutes les combinaisons possibles d'évènements successifs contenus dans une fenêtre temporelle de taille fixe sont considérées comme étant des motifs potentiels. Plusieurs notions comme l'attente ou le contexte temporel sont prises en compte par le modèle. Cependant, si assez prometteuse, cette méthode ne peut analyser de trop longues séquences car le coût algorithmique serait alors trop élevé. Par ailleurs, le problème de la polyphonie n'est pas vraiment considéré.

5.2.8 La similarité par partage de propriétés communes

Cambouropoulos [6] propose une méthode faisant partie d'une théorie générale d'analyse musicale que nous avons déjà décrite globalement au deuxième chapitre. Il propose en particulier un algorithme semi-automatique de découverte et de classification des motifs à partir de séquences monophoniques MIDI. Le rythme est supposé connu à l'exception de la métrique. Son algorithme s'articule en deux temps :

D'abord, la séquence est segmentée. Pour cela, une pondération de chacune des notes de la séquence est effectuée à partir de règles inspirées de la Gestalt. Une grille métrique est alors choisie de manière à correspondre le mieux possible à la pondération. Tous les motifs possibles contenant le même nombre de notes sont alors comparés. La valeur de similarité entre deux segments est proportionnelle au nombre de propriétés qui leur sont communes. En quelque sorte, l'algorithme recherche donc des appariements exacts à partir de différentes représentations obtenues à partir des propriétés. Deux séquences peuvent être similaires sans être identiques si seules quelques unes de leurs propriétés sont partagées. Les extrémités des motifs dont la similarité dépasse un seuil sont ensuite pondérées. Ces poids s'ajoutent aux poids attribués initialement à chacune des notes. Les poids les plus élevés

sont alors interprétés comme points de segmentation.

Dans un deuxième temps, les différents segments issus de l'étape précédente sont répartis en catégories d'après leur similarité. L'algorithme de mesure de similarité entre deux segments est le même qu'à la première étape. Un algorithme d'optimisation itérative de la catégorisation est proposé. L'optimisation se fait en variant les pondérations attribuées à chacune des propriétés d'un segment musical ce qui a pour effet de modifier les valeurs de similarité entre deux segments.

L'algorithme que propose Cambouropoulos articule de façon originale les notions de frontières avec celles de parallélisme. Par ailleurs, la recherche de structures se fait sans a priori sur la forme, et les heuristiques de l'algorithme sont présentées comme étant les plus générales possibles. Cependant, les détails de la procédure de découverte de motifs ne semblent pas tous pertinents. Par exemple, la comparaison de tous les motifs possibles précédant la segmentation de la séquence induit un temps de calcul très long. Pour le réduire, les solutions proposées consistent par exemple à limiter la taille des patterns comparés, forcer leurs extrémités à correspondre avec des événements fortement pondérés ou utiliser une propriété pour faire un premier filtrage de l'espace des solutions. Ces différentes contraintes rendent l'algorithme assez rigide puisqu'elles réduisent toutes a priori l'espace des solutions possibles. Certains motifs que l'on jugerait perceptivement pertinents pourraient donc ne pas intervenir pour la segmentation de la séquence basée sur le parallélisme. Par ailleurs, le fait de comparer par une opération booléenne les propriétés des motifs, et pas les motifs eux-mêmes rend la mesure de similarité assez peu précise. Des motifs perceptivement similaires mais dont toutes les propriétés sont légèrement différentes ne seront pas détectés.

5.2.9 Synthèse

Construire un motif progressivement ?

Parmi les différents algorithmes qui ont été décrits, ceux qui s'avèrent les plus convaincants font reposer leur système d'extraction de motifs sur une analyse des similarités entre les différents segments contenus dans la séquence à analyser.

Les méthodes diffèrent aussi par la manière d'analyser les similarités. Deux types d'approches différents sont proposés. Le premier, progressif, induit des motifs de plus en plus grands à mesure que la séquence est parcourue. Les problèmes de mesure de similarité et d'extraction de motifs sont couplés. Le deuxième type d'approche, par étapes ou modules distincts, commence par définir une mesure de similarité entre deux séquences, puis propose un découpage de la séquence entière en fonction des différentes groupes de segments détectés par cette mesure.

La première approche s'avère être très complexe car elle englobe en un seul processus tous les aspects du problème. Elle permet de prendre en compte le contexte pour la reconnaissance des similarités. Elle est aussi peut-être plus intuitive, car on peut penser que notre perception des similarités entre motifs à l'écoute d'une œuvre se fait progressivement, alors que la modélisation de la deuxième approche opère des similarités par blocs. Cependant, ça n'est pas prouvé. On pourrait par exemple évoquer le cas du motif qui nous apparaît subitement comme un tout seulement après l'audition de son dernier élément. C'est alors l'audition de la totalité du motif qui est nécessaire pour sa reconnaissance, et non sa construction par accumulation de fragments. Cela confirmerait le fait que le motif possède une unité qui lui est propre et qui le fait se détacher des autres événements de la séquence. Cette hypothèse ferait davantage peser la balance pour une utilisation de la deuxième approche. Par ailleurs, la deuxième approche, par son découpage du problème en deux modules, s'avère moins complexe à modéliser, même si tout comme pour la première approche, les algorithmes décrits précédemment n'ont pas encore

pu être appliqués de façon satisfaisante à la musique polyphonique du fait du grand saut dans la complexité que cela représente par rapport à la musique monophonique.

Nous allons donc adopter la deuxième approche pour notre modèle, et commencer dans un premier temps par construire une mesure de similarité entre deux séquences données. Une hypothèse sera que la similarité entre deux séquences peut être modélisée sans considération sur le contexte, c'est à dire sans considération sur les événements ayant précédé ceux qui sont comparés. Cette hypothèse peut être vue comme une simplification nécessaire dans un premier temps pour pouvoir aborder le problème de la polyphonie. Elle pourra être remise en question dans un deuxième temps.

Des contraintes pour réduire l'explosion combinatoire

Un des grands problèmes rencontrés par les algorithmes d'extraction de motifs consiste à réduire l'explosion combinatoire relative au nombre potentiel de motifs pouvant être trouvés. Une solution peut être de considérer un ensemble de règles limitant le nombre de motifs possibles, mais ces règles sont très dures à définir si l'on veut qu'elles s'appliquent d'une façon générale à l'ensemble des cas de figure possibles. Souvent une règle va avoir du sens pour un morceau donné, mais sera contredite lorsqu'on cherchera à l'appliquer à d'autres morceaux. C'est le cas par exemple de l'extrait donné figure 5.5.

Il faut donc essayer de minimiser le nombre de règles. Une autre solution consiste à simplifier la partition, c'est à dire à diminuer le nombre de notes et par là limiter l'explosion combinatoire. Par exemple, les notes considérées comme non importantes pourraient être supprimées. Nous avons choisi pour aller dans ce sens une importante réduction du problème en ne considérant que les événements (notes et silences) situés sur les pulsations pour certaines de nos mesures de similarité (nous en définirons trois différentes et complémentaires). Cela permet à la fois de réduire la complexité combinatoire, mais aussi d'autoriser certaines variations dans les séquences à comparer du fait que les événements situés entre les pulsations pourront différer sans altérer la



FIG. 5.5: Les deux premières séquences seq1 et seq2 sont perceptivement assez similaires. Elles pourraient répondre à la règle : deux séquences de même taille dont les intervalles entre notes successives ne diffèrent pas plus de un ton sont similaires. Pourtant, la séquence seq3 comparée à seq1 respecte cette règle mais est peu similaire à seq1.

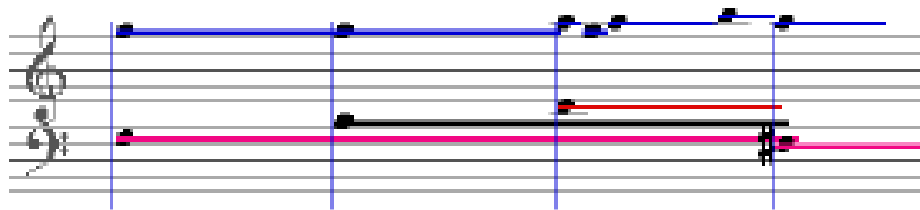


FIG. 5.6: Une séquence de segments de pulsations extraite des *Variations Goldberg* de Bach. Les barres verticales délimitent les segments. Les barres horizontales représentent les durées de chaque évènement.

mesure de similarité. Un autre argument est que les évènements situés sur les pulsations sont souvent perçus avec plus d'acuité et donc ont un poids perceptif potentiellement plus grand que les autres évènements, ce qui fait que si certains évènements devaient être conservés, ce devrait être ceux situés sur les pulsations. Dans nos exemples, nous symboliserons par des barres verticales les emplacements des évènements situés sur les pulsations. Un exemple est donné figure 5.6.

5.3 Extraction des épures de deux séquences polyphoniques

La mesure de similarité la plus simple à modéliser est celle qui repère les répétitions exactes. Si ce cas de figure peut sembler trivial, il n'en est pas moins très souvent utilisé en musique, à différents niveaux. Par exemple, un thème ou une phrase peuvent souvent être immédiatement répétés. Plus généralement, les répétitions peuvent faire partie intégrante d'une forme musicale, par exemple les expositions d'une sonate, le refrain d'un rondo, ou la reprise du menuet dans la forme menuet.

Si cette mesure de similarité ne présente pas en elle-même de difficulté conceptuelle à nos yeux, elle n'en reste pas moins difficile à modéliser algo-

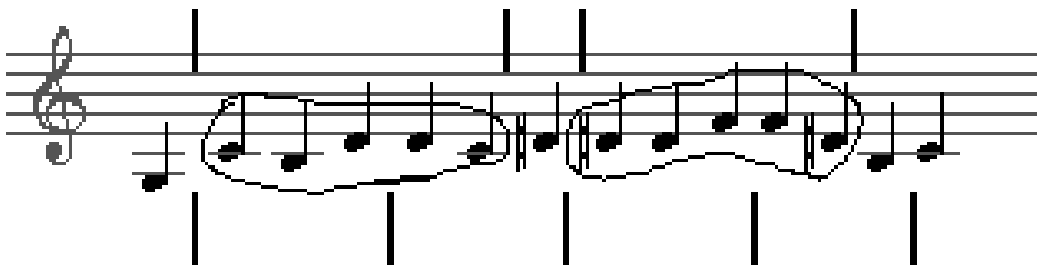


FIG. 5.7: Une séquence segmentée de deux façons différentes : la première segmentation (haut de la figure) a été faite suivant les notes correspondant à la plus grande répétition exacte. La deuxième segmentation (bas de la figure) a été faite à l'oreille. La deuxième segmentation semble plus satisfaisante.

rhythmiquement sous l'angle de la perception. En effet, les séquences perçues comme identiquement répétées ne sont pas toujours celles que l'on croit. La figure 5.7 illustre l'exemple d'une séquence dont l'analyse ne serait pas très satisfaisante si le découpage se faisait selon les répétitions exactes des plus grandes sous-séquences.

Considérons maintenant le problème de la mesure de similarité entre deux séquences polyphoniques quelconques.

5.3.1 Une notion de départ : l'intersection

Dans une première approximation, pour mesurer la similarité entre deux séquences différentes, un algorithme assez simple consisterait à calculer l'intersection entre les deux séquences, de façon à repérer les éléments communs. On pourrait de cette manière détecter par exemple les répétitions exactes de mélodies ayant un accompagnement différent. L'intersection pourrait aussi porter non pas sur les notes, mais comme le propose Cambouropoulos [6] sur différentes propriétés globales de la séquence, comme par exemple le tempo, le contour, le nombre de notes ou la suite des hauteurs. La similarité serait alors proportionnelle au nombre de propriétés partagées, une propriété ne

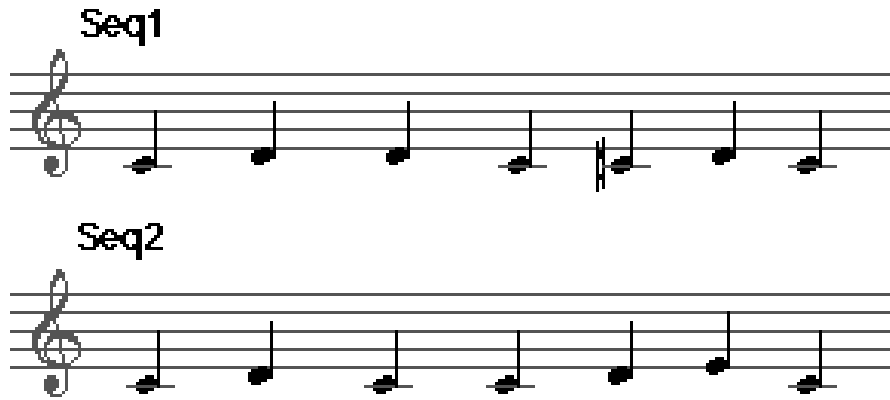


FIG. 5.8: Deux séquences localement différentes (les intervalles diffèrent localement de 0 à 1 ton), mais globalement similaires.

pouvant être à moitié partagée.

Cependant, il serait facile d'imaginer d'autres exemples de séquences perceptivement similaires mais n'ayant aucun élément ou propriété identiques. Il suffirait par exemple que les séquences diffèrent localement tout en étant globalement perçues similaires (mais pas identiques), comme c'est le cas pour les séquences seq1 et seq2 représentées figure 5.5.

Faut-il pour autant chercher une autre méthode ? Pas vraiment, car cette dernière reste assez proche de notre propre façon d'établir les similarités. En effet, nous avons tendance à rechercher davantage les éléments communs à deux séquences plutôt que les éléments différents, ce qui revient en quelque sorte à effectuer l'intersection des deux séquences. Seulement, notre perception autorise les éléments de l'intersection à être différents du moment qu'ils restent similaires. Si nous voulons utiliser cet algorithme, il nous faut donc l'adapter au phénomène musical, ce qui revient à définir un nouvel algorithme proche de l'intersection mais faisant intervenir la notion de similarité.

Nous définissons la similarité entre la séquence x et la séquence y (x est d'abord écoutée, donc mémorisée, puis comparée à y) par la proportion du nombre d'évènements contenus dans x qui sont similaires à certains évènements de y . Inversement, la similarité entre y et x dépendra du nombre

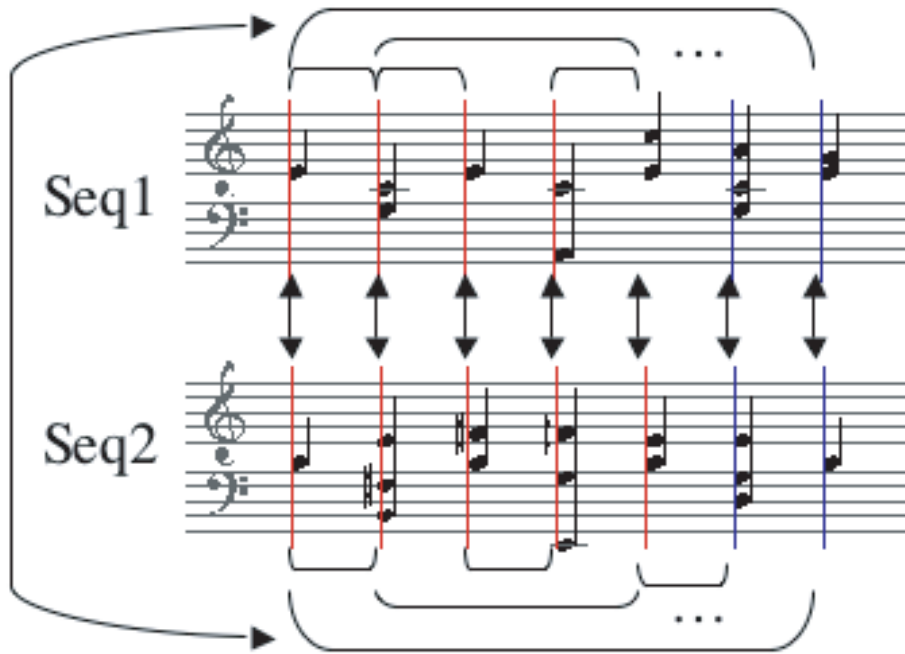


FIG. 5.9: Deux séquences de même taille (en nombre de pulsations) sont comparées. A chaque position de pulsation d'une séquence correspond une et une seule position de pulsation dans l'autre séquence. La similarité se détermine alors en comparant les événements et les intervalles situés ou définis par des positions se correspondant

d'évènements contenus dans y qui seront similaires à certains évènements de x . Notons la non-symétrie de la relation, qui sera précisée par l'équation 5.2 de la partie intitulée 'les règles cognitives'. Un exemple est donné figure 5.10.

Puisque nous avons supposé que les séquences comparées avaient même nombre de pulsations, il est possible d'associer à chaque position de pulsation d'une séquence une et une seule position correspondante de l'autre séquence. De même, un intervalle d'une séquence pris entre deux notes situées aux positions pulsation2 et pulsation5 peut être associé à tous les intervalles de l'autre séquence pris entre les notes des positions pulsation2 et pulsation5. L'algorithme de similarité entre deux séquences consistera donc de façon générale à comparer à la fois les évènements dont les positions se correspondent



FIG. 5.10: Deux séquences similaires extraites de *l'art de la fugue* de Bach. Seq1 n'est véritablement similaire qu'avec les éléments surlignés de seq2, mais les deux séquences entières auront la même valeur de similarité puisque tout les éléments de seq1 participent de la similarité avec une partie de seq2. Par contre, la similarité entre seq2 et seq1 (seq2 étant cette fois pris comme référence) sera différente puisque seule une partie de la séquence de référence seq2 intervient dans la similarité.

et les intervalles dont les positions des notes les définissant se correspondent. Un exemple est donné figure 5.9.

Nous allons maintenant définir les types de similarité que peuvent entretenir les évènements des séquences comparées.

Pour cela, nous avons décidé de prendre en compte trois aspects qui correspondent selon nous à trois dimensions musicales très importantes :

- le contour des hauteurs ;
- les intervalles de hauteurs ;
- le rythme.

Ce choix, s'il peut paraître arbitraire dans un premier temps, repose tout de même sur des points de vue depuis longtemps utilisés en analyse musicale qui nous semblent alors incontournables même s'ils pourraient être associés à d'autres paramètres musicaux, comme par exemple la dynamique. Le rythme est, nous l'avons vu au chapitre trois, indispensable à la musique. Le contour de hauteurs est l'une des notions la plus unanimement reconnue et partagée, autant par les auditeurs que par les compositeurs ou les analystes. Nous verrons que cette notion reste cependant difficile à modéliser, un contour ne

pouvant se résumer à une suite de variations d’intervalles entre des notes successives extrêmes d’une polyphonie. Les intervalles de hauteurs sont plus difficiles à définir. Ils concernent aussi bien les intervalles verticaux d’une polyphonie, c’est à dire les accords, que les suites d’intervalles horizontaux donc les mélodies incluses dans la polyphonie.

Avant de décrire notre modèle de similarité selon chaque point de vue, nous allons définir quelques principes cognitifs généraux qui peuvent être vus comme des règles à appliquer aux mesures pour qu’elles aient des chances d’avoir une cohérence perceptive.

5.3.2 Les règles cognitives

La première règle est que chaque séquence musicale doit être considérée par chaque mesure comme une possible entité formant un tout, c’est à dire davantage qu’une simple concaténation de plus petites séquences. Cela implique que la mesure de similarité ne soit pas linéaire :

$$S(x, x') \text{ et } S(y, y') \not\propto S(xy, x'y') \quad (5.1)$$

où $\propto$ signifie ‘ne nous informe pas sur’, où $S(x, x')$ représente la fonction de similarité entre les séquences x et x' et où xy désigne la séquence issue de la concaténation entre x et y .

En effet, la mesure de similarité dépend des relations à la fois locales et globales entre les éléments. Définir une mesure linéaire reviendrait à ne considérer que les relations locales. Ce cas est illustré par les figures 5.5 et 5.11.

Une deuxième règle est la non symétrie de la mesure de similarité :

$$S(x, x') \not\propto S(x', x) \quad (5.2)$$

En effet, en supposant que x soit similaire à x' , donc potentiellement similaire avec une sous-séquence de x' , la relation de symétrie impliquerait que x' ait la même valeur de similarité avec une sous-séquence de x . Or, ce n’est pas

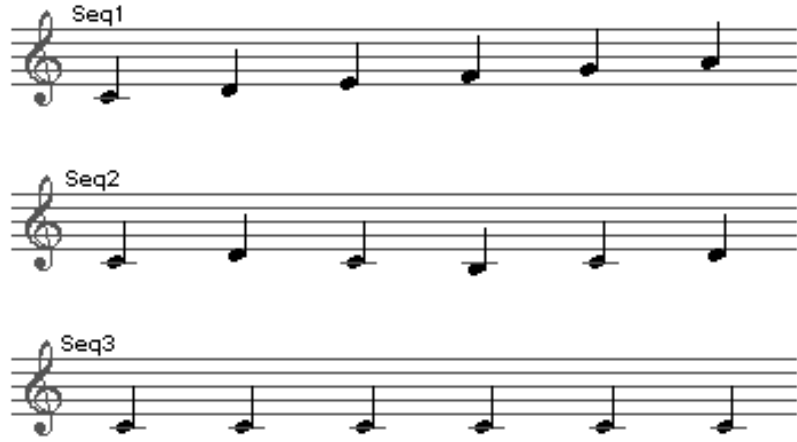


FIG. 5.11: Seq1 et Seq3 ont des intervalles consécutifs similaires, mais les deux séquences sont globalement perçectivement différentes. Seq2 et Seq3 sont à la fois localement et globalement similaires. Les intervalles non consécutifs doivent être pris en compte dans la mesure de similarité qui n'est donc plus linéaire.

toujours le cas, comme l'illustre par exemple la figure 5.10. De plus, si l'on recherchait indifféremment l'inclusion de x dans x' ou x' dans x , on pourrait trouver par exemple que la première moitié de x est contenue, donc similaire à la première moitié de x' , et que la deuxième moitié de x' est contenue donc similaire à la deuxième moitié de x , et en conclure par symétrie, et en supposant que la règle1 ne soit pas respectée, que la séquence entière x est similaire à x' . Un tel cas se présenterait par exemple pour les deux séquences représentées figure 5.12.

Enfin, une troisième règle concerne la non transitivité de la mesure :

$$S(x, y) \text{ et } S(y, z) \not\propto S(x, z) \quad (5.3)$$

Par exemple, z peut être une variation de y qui est lui-même une variation de x . Mais z peut être très différent de x et ne pas respecter l'inégalité triangulaire.

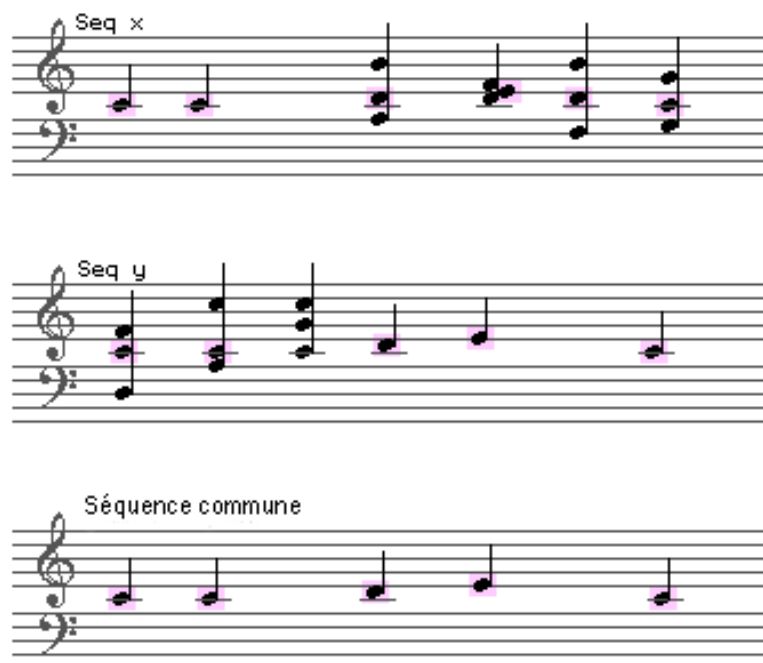


FIG. 5.12: Les séquences x et y pourraient à première vue être considérées similaires, puisque leur intersection n'est pas vide. Le début de x est inclus dans y et la fin de y est incluse dans x, mais les séquences ne sont pas globalement similaires parce qu'aucune des deux séquences n'est entièrement incluse dans l'autre.

Ces trois principes constituent pour nous une base de règles perceptives qui devront être respectées par les différents types de similarité que nous allons maintenant expliciter.

5.3.3 La similarité par contour de hauteurs

Dans beaucoup de séquences musicales, le contour des hauteurs apparaît comme perceptivement plus saillant que les autres séquences d'intervalles. Nous allons essayer de le modéliser, en commençant par définir ce qu'est pour nous un contour.

Qu'est-ce qu'un contour de hauteurs ?

Un contour de hauteurs est souvent représenté par les variations des hauteurs situées à l'extrémité supérieure d'une séquence musicale (un exemple est donné figure 5.13). D'autres représentations sont possibles, par exemple pour affiner la représentation proportionnellement à l'intensité des variations de hauteurs. On pourrait alors définir deux types de contour vers le bas (- et - -) et deux types de contours vers le haut (+ et ++).

Pour établir une valeur de similarité entre deux séquences, une telle représentation ne permettrait pas d'aboutir à une mesure très fine. D'une part, la représentation est figée quelles que soient les séquences à comparer. Dans le cas du contour simple -, 0, +, des mélodies assez différentes pourraient avoir le même contour, et donc la même valeur de similarité que des mélodies ayant un contour beaucoup plus proche. Dans le cas du contour affiné, des frontières absolues devraient être définies pour les différents types de variations ce qui peut aboutir à des résultats incohérents. Par exemple, la figure 5.14 montre deux représentations de contours différents mais qui seraient jugés perceptivement similaires. Les intervalles de plus d'une tierce sont notés ++ ou - et les autres + ou -.

D'autre part, dans le cas d'une polyphonie, les contours perçus ne correspondent pas forcément à la suite d'intervalles entre chaque évènement

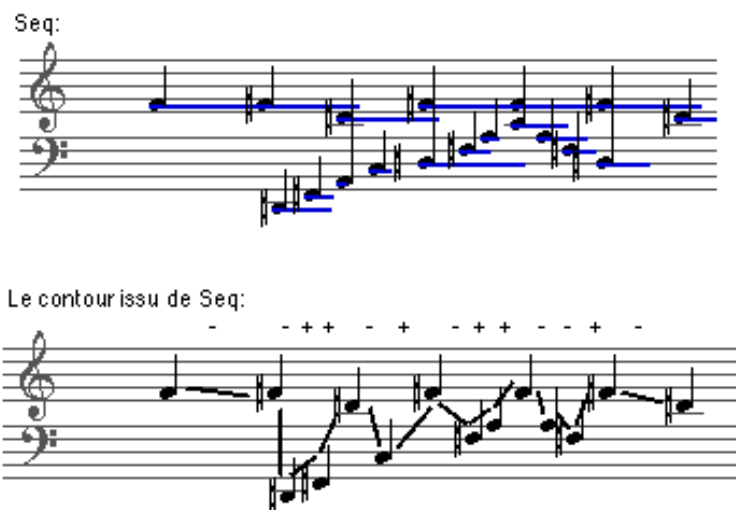


FIG. 5.13: Extrait de l'*Intermezzo* op118 no6 de Brahms. Le contour est défini par les variations des intervalles entre les notes consécutives de la polyphonie. Cette représentation mélange l'accompagnement et la voix du haut, et n'est donc pas satisfaisante.



FIG. 5.14: Deux séquences aux contours perceptivement assez proches. Cependant, la notation choisie ne permet pas d'en rendre compte puisqu'elle représente différemment les deux contours. Les intervalles de plus d'une tierce sont notés ++ ou -- et les autres + ou -. La représentation par contours affinés ne convient donc pas pour cet exemple.

successif. Par exemple, la figure 5.13 montre un contour mélangeant la mélodie supérieure avec l'accompagnement. Il nous faut donc établir un contour qui soit à la fois plus proche de notre perception et qui permette d'établir des mesures de similarité plus précises.

Nous définissons un contour de hauteurs comme étant la suite des valeurs perceptivement extrêmes d'une séquence de hauteurs. Le contour peut donc se situer aux parties supérieures de la suite de hauteurs, mais aussi aux parties inférieures. En effet, la partie basse d'une polyphonie, comme la partie haute mais dans une moindre mesure, est souvent perçue davantage que les parties intermédiaires.

Par ailleurs, le contour n'est plus défini par les variations de hauteurs, mais par les hauteurs elles-mêmes. Les variations seront prises en compte différemment selon chaque cas particulier de comparaison entre deux séquences.

Enfin, le contour se détermine suivant les valeurs perçues comme extrêmes, mais pas forcément extrêmes au sens mathématique. Par exemple, une mélodie peut être perçue même si des notes d'accompagnement se trouvent insérées entre ses éléments.

Pour modéliser ces valeurs perceptives nous proposons de considérer à la fois les hauteurs et les durées. Les hauteurs faisant partie alors du contour sont celles par exemple des parties supérieures qui ne sont pas recouvertes par une durée provenant d'une hauteur supérieure à elles. La durée est donc l'un des paramètres qui est utilisé pour nous faire percevoir la continuité entre les notes du contour. De nombreux autres paramètres tels que l'accent dynamique, le timbre ou le staccato peuvent aussi être utilisés, mais nous ne les modéliserons pas dans un premier temps. La figure 5.15 propose un exemple d'une telle représentation.

La mesure de similarité entre deux contours de hauteurs

Chacun des deux contours d'une séquence est comparé aux deux contours de l'autre séquence. Puisque les deux séquences ont le même nombre d'éléments (en nombre de pulsations), chaque valeur d'un contour peut être as-

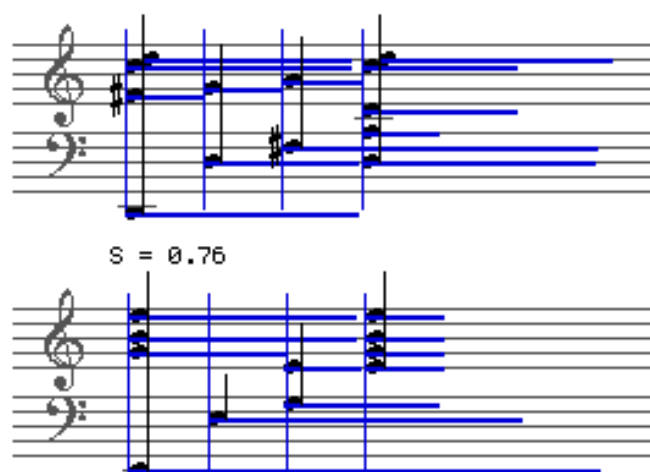


FIG. 5.15: Deux séquences similaires issues de l'intermezzo op119 n°3 de Brahms. Les contours du haut (ré-ré et mi-mi) sont similaires. Les hauteurs situées entre les notes extrêmes des deux séquences ne sont pas incluses dans les contours parce que les durées (lignes horizontales) des premières hauteurs (ré et mi) de chaque contour les recouvrent.

sociée à une valeur de l'autre contour. Nous définissons alors la similarité entre deux contours à partir de la comparaison de toutes les paires d'intervalles possibles entre chaque valeur de contour se correspondant. Nous considérons donc à la fois les intervalles entre notes consécutives, mais aussi les autres intervalles, ce qui respecte la règle cognitive 5.3 précédemment définie. La figure 5.16 montre deux contours de 5 notes. Tous les intervalles des deux séquences se correspondant seront comparés. Pour donner plus d'importance aux intervalles consécutifs, la similarité de contour est définie par une moyenne entre la valeur de similarité moyenne obtenue pour les seuls intervalles consécutifs et la valeur moyenne obtenue pour tous les intervalles possibles.

La comparaison entre deux intervalles se correspondant consiste à donner un poids de similarité à la paire d'intervalles :

Si les intervalles sont égaux :

1 lorsque les intervalles ne sont pas dans un contour bas ;



FIG. 5.16: Deux contours de 5 notes. Tous les intervalles se correspondant ont été tracés. Chaque intervalle d'une séquence correspond à un intervalle de l'autre séquence ce qui permet de calculer la similarité entre chacun des intervalles se correspondant.

0.7 lorsque les intervalles sont dans un contour bas.

Si les intervalles sont de même signe :

1 lorsque les intervalles sont dans un contour haut et que
 $\text{abs}(\text{int1} - \text{int2}) < 4$;

0.6 lorsque les intervalles ne sont pas dans un contour haut et
que $\text{abs}(\text{int1} - \text{int2}) < 4$;

0.3 si $\text{abs}(\text{int1} - \text{int2}) < 10$;

0.1 sinon.

Si l'un des intervalles vaut 0 :

0.4 si $\text{abs}(\text{int1} - \text{int2}) < 3$;

0 si $\text{abs}(\text{int1} - \text{int2}) \geq 3$.

Si les intervalles sont de signes opposés :

0 .

Remarquons que des variations nulles d'intervalles comparées à des variations non nulles peuvent contribuer à la similarité entre les deux contours.

5.3.4 La similarité par intervalles de hauteur horizontaux

Par le terme intervalles de hauteur horizontaux, nous désignons les différentes lignes mélodiques contenues dans la polyphonie.

Comme pour les contours, nous analysons les séquences constituées des seuls évènements situés sur les battues de la pulsation.

Le problème

Etablir la similarité entre les différentes lignes mélodiques de deux polyphonies est un problème très complexe lorsqu'on a pour objectif que cette similarité soit perceptive. En effet, contrairement au contour qui se borne aux mélodies supérieures ou inférieures, les différentes lignes mélodiques contenues dans la polyphonie sont potentiellement très nombreuses et leur extraction résulte d'une combinatoire complexe. Si l'on fait abstraction des critères perceptifs, chaque séquence de hauteurs extraite de l'une des polyphonies est potentiellement similaire à n'importe quelle autre séquence extraite de l'autre polyphonie. Par ailleurs, si l'on raisonne de la même façon que pour les contours, deux lignes non identiques pourraient être jugées similaires par l'algorithme d'après les variations de leurs intervalles. Une exemple est donné figure 5.17.

Pour éviter l'explosion combinatoire, il faut soit isoler de la polyphonie les mélodies qui ont le plus de chances d'être perceptivement saillantes, donc importantes pour établir la similarité, soit établir des contraintes perceptives limitant les possibilités de comparaison entre les mélodies. Par exemple, une contrainte pourrait interdire la comparaison de deux lignes de contours différents.

La première solution est difficile à mettre en œuvre. Déjà, les mélodies situées aux extrémités d'une polyphonie sont difficiles à extraire. Lorsque les mélodies se situent à l'intérieur de la polyphonie, le problème se complique encore davantage. Par exemple, la figure 5.18 montre une polyphonie

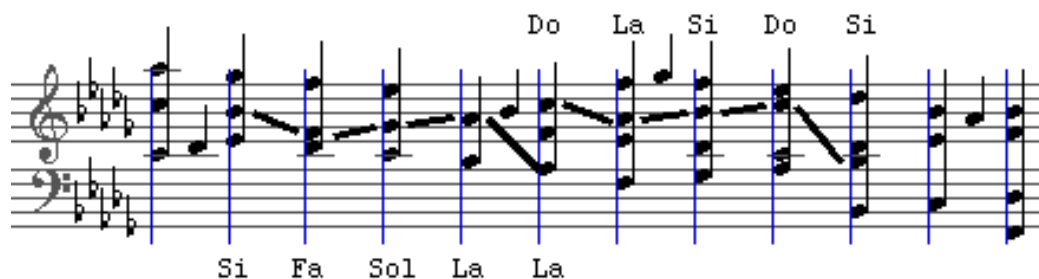


FIG. 5.17: Deux lignes mélodiques tirées de la fugue en Si bémol mineur du *Clavier bien tempéré* de Bach. Les intervalles se correspondant des deux lignes ne diffèrent pas de plus d'un ton. Pourtant, ces lignes ne seront sans doute pas perçues comme similaires à l'écoute du morceau.

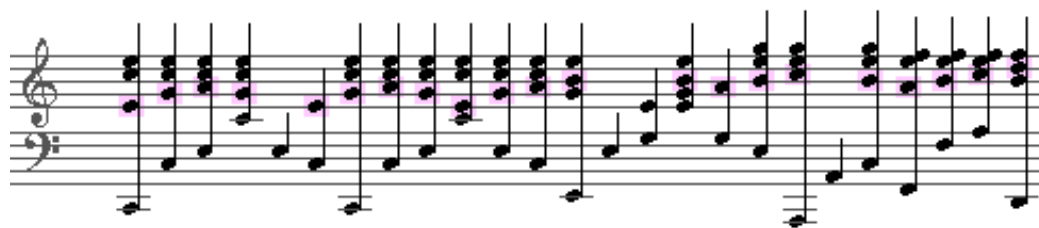


FIG. 5.18: Extrait de l'*Intermezzo* op119 no3 de Brahms. La mélodie perçue est surlignée. Elle se trouve au milieu de la polyphonie, et serait très difficile à extraire sans considérer les événements passés et futurs de la séquence.

à l'intérieur de laquelle la mélodie perçue se trouve presque cachée. Il paraît très difficile d'imaginer un algorithme qui pourrait l'extraire indépendamment des événements suivants ou précédents, c'est à dire indépendamment des répétitions. Nous avons donc choisi la deuxième solution, mais là encore, les combinaisons possibles sont très nombreuses dès lors que des différences sont autorisées entre les différentes mélodies. Par ailleurs, il est difficile de déterminer si après avoir isolé deux lignes non identiques mais similaires, elles seront toujours perçues comme telles une fois replacées à l'intérieur des deux polyphonies (nous avons déjà illustré ce cas par la figure 5.17). Nous avons donc décidé de rechercher uniquement les lignes de hauteurs identiques entre elles à quelques variations près, plutôt que de rechercher les lignes similaires. Cela correspond à l'hypothèse perceptive qu'une paire de séquences identiques a plus de chance d'être perçue comme telle que d'autres paires possibles de séquences similaires. De plus, la seule considération de l'identité limite agréablement notre espace de recherche. Sur l'ensemble de la polyphonie, la notion de similarité n'est pas pour autant remise en question. La mesure globale restera une mesure de similarité, puisque les différentes lignes de deux polyphonies similaires ne seront pas toutes identiques entre elles, même si au niveau d'une ligne les mesures porteront seulement sur l'identité. De plus, notre choix respecte les différents paramètres cognitifs que nous avons définis dans la section précédente. En particulier, nous allons rechercher des lignes de hauteur continûment identiques sur l'ensemble de la séquence, de façon à prendre en compte son unité, plutôt que d'agréger des segments discontinus de lignes localement identiques. Un exemple de séquences localement similaires mais globalement différentes est donné figure 5.19.

L'algorithme d'extraction de lignes identiques

L'identité est comprise ici à une transposition près : deux lignes de hauteurs sont identiques si une transposition de l'une permet d'obtenir l'autre.

Etant donné deux séquences polyphoniques *seq1* et *seq2*, notre objectif est de trouver les lignes de hauteurs de *seq2* qui sont identiques (à une trans-

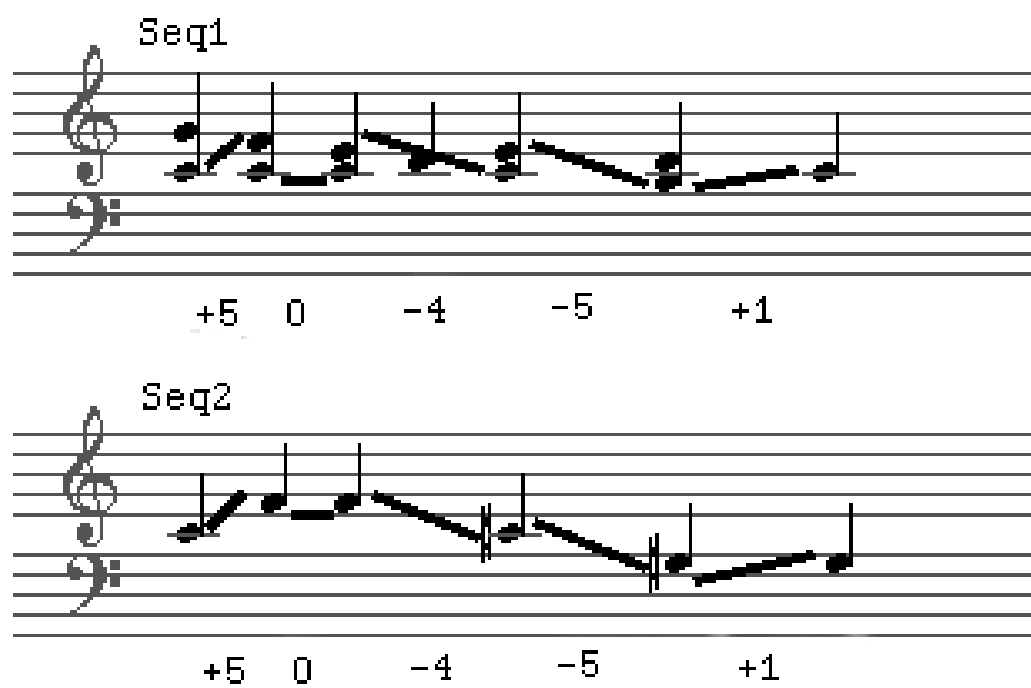


FIG. 5.19: Seq2 partage localement les mêmes intervalles que Seq1. Pourtant, les deux séquences sont globalement perceptivement différentes.

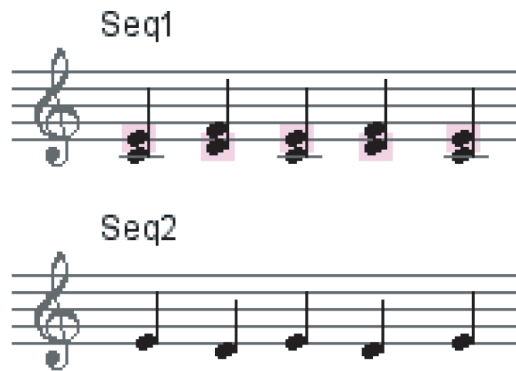


FIG. 5.20: Deux séquences contenant une même mélodie.

position près) à une ligne de hauteur de $seq1$. Par exemple, la figure 5.20 montre deux séquences différentes contenant une même mélodie transposée. Cet exemple illustre la difficulté d'extraire une mélodie qui fasse perceptivement sens, puisque même lorsque deux mélodies sont identiques, il est possible que d'autres événements de la polyphonie interfèrent dans leur perception.

Pour simplifier la représentation du modèle, nous supposons qu'à chaque position de pulsation de chaque séquence correspond un accord. Nous comparons donc deux séquences ayant chacune le même nombre d'accords L . Notre algorithme effectue alors $L-1$ itérations. Chaque itération i compare les séquences $seq1$ et $seq2$ dont les premiers $i-1$ accords ont été tronqués. La même fonction de comparaison est employée pour chaque itération i . Cette fonction a pour sortie une valeur de similarité pouvant varier de 0 à 1. Au total, la similarité entre les deux séquences sera calculée en faisant la moyenne des similarités obtenues à chacune des $L-1$ itérations.

Nous allons maintenant décrire la fonction de comparaison appliquée à la première itération. A partir d'une hauteur du premier accord de $seq1$ que nous appelons $hauteur(i)(accord1(seq1))$, nous effectuons autant de transpositions de $seq2$ qu'il y a de notes contenues dans son premier accord. La transposition consiste à ajouter aux hauteurs de $seq2$ une constante telle que l'une des

hauteurs du premier accord ($\text{accord1}(\text{seq2})$) soit égale à la hauteur de seq1 que nous avons considérée $\text{hauteur}(i)(\text{accord1}(\text{seq1}))$. La taille en nombre d'accords (de valeur maximale $L-i$) de la plus grande intersection entre seq1 et les différentes transpositions de seq2 est à chaque fois mémorisée. Cette taille est ensuite moyennée par la taille totale L des séquences. Le fonction de comparaison s'applique ainsi aux N hauteurs du premier accord de seq1 , ce qui fournit N moyennes comprises entre 0 et 1. Le résultat final de la fonction de comparaison appliquée à la première itération est la moyenne entre ces N valeurs.

La fonction de comparaison est ensuite appliquée de la même façon pour chaque itération i sur les séquences seq1 et seq2 dont les $i-1$ premières valeurs ont été tronquées.

5.3.5 La similarité par intervalles de hauteur verticaux

Par le terme intervalles de hauteur verticaux, nous voulons désigner les différents accords contenus dans la polyphonie. La comparaison des accords nous semble être un critère important pour établir la similarité entre deux séquences polyphoniques.

Chaque accord est représenté par la liste des intervalles entre la basse et les autres notes. Les intervalles sont considérés modulo 12. Cette représentation différencie un accord et ses renversements. Pour ne pas faire cette différence, on pourrait utiliser d'autres représentations, par exemple celle que propose la set-theory [18], mais se poserait alors la question de la généralité de l'algorithme, puisque par exemple la set-theory considère non seulement les renversements d'accords comme équivalent, mais aussi les symétries par rapport à la note fondamentale ce qui ne fait pas toujours perceptivement sens.

Nous avons ensuite considéré deux aspects :

- la similarité d'accord entre chaque séquence ;
- la similarité de répétition d'accord entre chaque séquence.

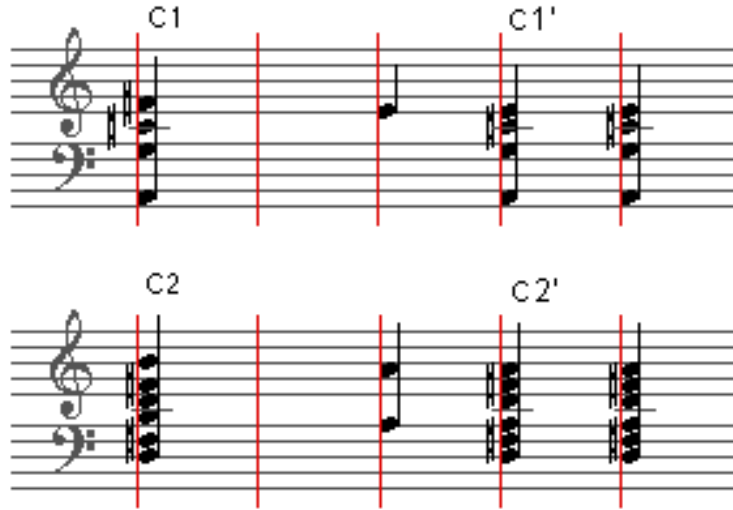


FIG. 5.21: Extrait de la *sonate* en RéM d664 op121 deuxième mouvement de Schubert. Les accords C1 et C2 sont similaires (mêmes intervalles de tierce majeure et de sixte mineure à partir de la basse) ainsi que C1' et C2' (intervalles de tierce majeure et quinte à partir de la basse).

La similarité d'accord entre chaque séquence

Nous supposons que deux accords similaires dont les positions dans chacune des séquences se correspondent, seront perçus comme tels par un auditeur.

La similarité entre deux accords C.1 et C.2 de taille *taille1* et *taille2* (en nombre d'intervalles) est alors définie par :

$$S_{accord}(C.1, C.2) = \frac{taille\ C.1 \cap\ C.2}{\min(taille1, taille2)} \quad (5.4)$$

Par exemple, la figure 5.21 montre deux séquences d'accord similaires. La similarité globale repose pour une grande partie sur la similarité entre les paires d'accords se correspondant.

La similarité de répétition d'accord entre chaque séquence

Le second aspect que nous avons pris en compte concerne les similarités de répétition d'accord entre chaque séquence. L'idée est que si deux accords, même différents, sont répétés dans chaque séquence sur des positions correspondantes, une similarité entre les deux séquences sera perçue.

Plus formellement, la similarité entre deux paires d'accords se correspondant $C.1$, $C.1'$ et $C.2$, $C.2'$ (voir figure 5.22) de tailles $taille1$, $taille1'$, $taille2$ et $taille2'$ est définie par :

$$S_{repet}((C.1, C.1'), (C.2, C.2')) = \min\left(\frac{\text{taille } C.1 \cap C.1'}{\min(\text{taille1}, \text{taille1'})}, \frac{\text{taille } C.2 \cap C.2'}{\min(\text{taille2}, \text{taille2'})}\right) \quad (5.5)$$

Nous avons pris le minimum de chaque comparaison d'une paire d'accords car si l'une des séquences n'a pas de répétition d'accord aucune similarité ne pourra être établie entre les deux séquences, même si l'autre séquence contient une répétition d'accord.

Par exemple, la figure 5.22 montre deux séquences similaires d'accords. Cette similarité repose pour une grande partie sur la similarité de répétition d'accord entre chaque séquence, c'est à dire que les accords $C1$ sont répétés plusieurs fois ainsi que les accords $C2$ leur correspondant.

Regroupement des deux mesures de similarité d'accord

Globalement, pour prendre en compte les deux types de mesures de similarité d'accord que nous venons de définir, nous calculons une moyenne sur le premier type, puis nous prenons la valeur maximale entre les deux types :

$$S_{total}((C.1, C.1'), (C.2, C.2')) = \max(\text{mean}(S_{accord}(C.1, C.2), S_{accord}(C.1', C.2')), S_{repet}((C.1, C.1'), (C.2, C.2'))) \quad (5.6)$$

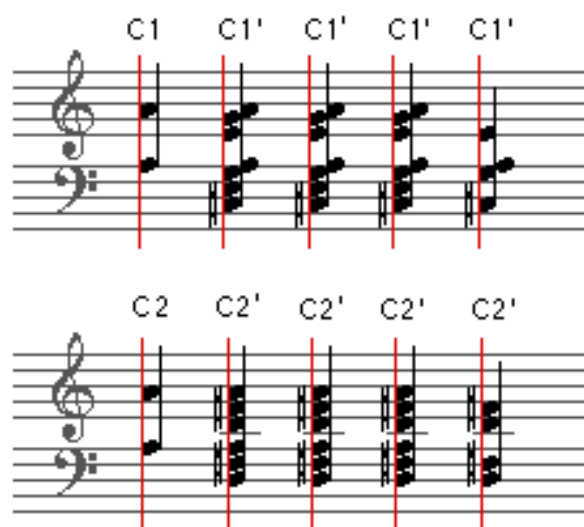


FIG. 5.22: Extrait de la *sonate* en RéM d664 op121 deuxième mouvement de Schubert. Les accords C1' et C2' sont différents, mais ils sont tous deux répétés plusieurs fois ce qui constitue une forme de similarité entre les deux séquences.

5.3.6 La similarité rythmique

Le rythme est un composant essentiel à la musique, mais il est assez rarement pris en compte par les algorithmes d'extraction de motifs. Une raison pourrait être que deux séquences mélodiques identiques dont on varie le rythme restent souvent perceptivement très similaires, alors que inversement deux séquences dont on varie les hauteurs sans modifier le rythme seront plus rapidement perçues comme différentes. Les hauteurs joueraient ainsi un rôle plus important que le rythme dans notre perception des similarités. Cependant, cet argument ne nous satisfait pas, et nous tenons le rythme comme un composant aussi essentiel que les hauteurs pour l'extraction des répétitions. Une autre raison plus pragmatique serait que dans un premier temps, on préfère souvent laisser le rythme au second plan pour n'étudier que les hauteurs, qui sont sans doute plus faciles à manipuler puisqu'elles peuvent facilement être extraites hors du contexte temporel. Quoi qu'il en soit, le rythme reste un paramètre important qui peut se révéler absolument nécessaire pour la détection de certains motifs.

Comme pour les mesures de similarité précédentes, notre modèle compare la structure rythmique de deux séquences polyphoniques de même taille. Cependant, contrairement aux mesures précédentes, tous les événements de la séquence sont comparés, et non pas seulement les événements situés sur les battues de pulsation. La figure 5.23 montre l'exemple de deux séquences très similaires, l'une des deux séquences étant rythmiquement incluse dans l'autre.

Dans une première étape, les séquences sont normalisées de manière à ce qu'elles aient le même tempo. Ensuite, les différents onsets des deux séquences sont appariés lorsque c'est possible, c'est à dire lorsque leurs valeurs sont assez proches. Pour cela, les segments de pulsation se correspondant sont comparés. Pour chaque onset contenu dans un segment de pulsation on calcule sa proportion par rapport à la taille du segment. Deux onsets de deux segments sont alors appariés si leurs proportions respectives par rapport au segment ne diffèrent pas en pourcentage de plus de 'approx-on'. Par

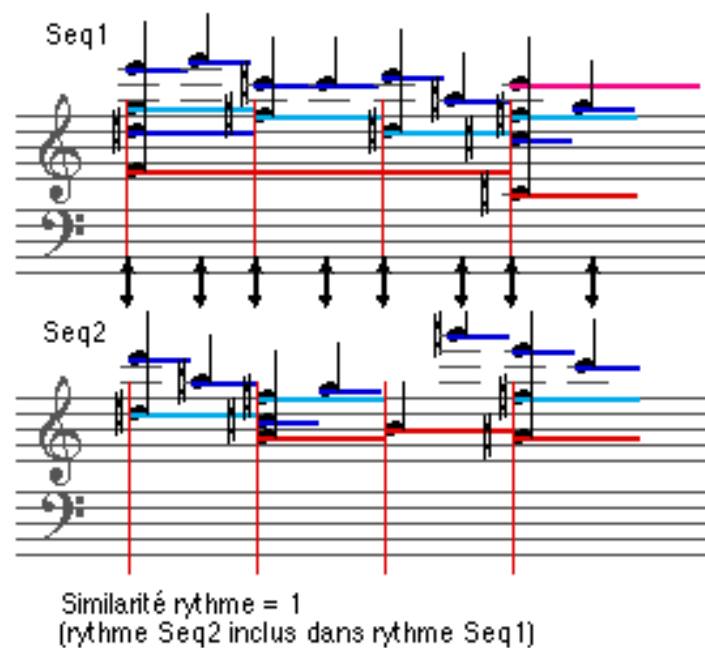


FIG. 5.23: Deux séquences rythmiques similaires issues du *Pierrot Lunaire* de Schoenberg. Les onsets se correspondant sont repérés par une flèche. La similarité rythmique est de 1 car le rythme de seq2 est inclus dans celui de seq1.

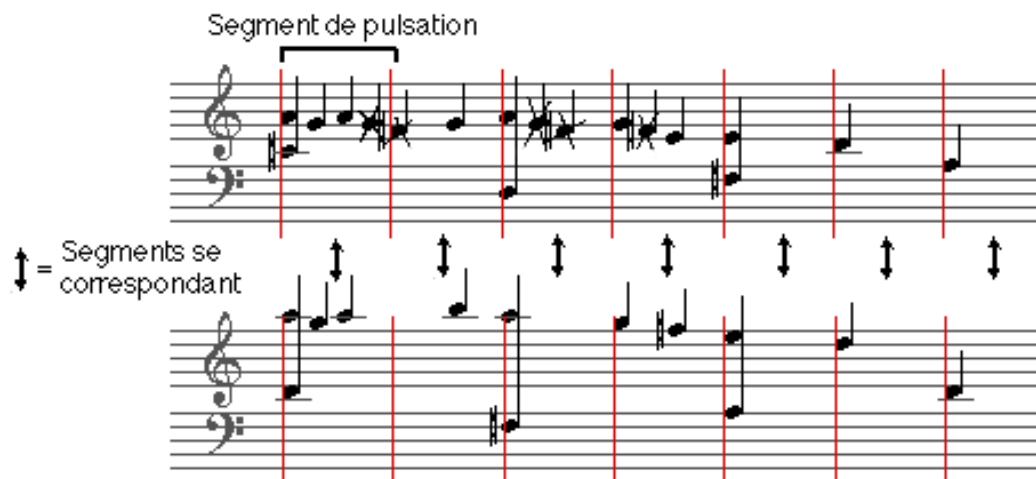


FIG. 5.24: Deux extraits de l'aria des *Variations Goldberg* de Bach. Les deux séquences ont chacune 7 segments de pulsation. Chaque segment d'une séquence peut être mis en correspondance avec un seul segment de l'autre séquence. Pour chaque couple de segments en correspondance, certains onsets sont éliminés s'il ne leur correspond pas d'onset similaire dans l'autre séquence (ils sont alors barrés par une croix sur la figure).

défaut, 'approx-on' est fixé à 0.1. Les onsets non appariés sont alors supprimés. L'exemple donné figure 5.24 montre deux séquences de segments de pulsation se correspondant. Pour chaque couple de segment, les onsets sont appariés s'ils ont un rapport de proportion proche.

La similarité entre deux séquences Seq1 et Seq2 est alors définie comme étant la moyenne de la similarité entre chaque paire de segment de pulsation. Prenons la séquence Seq1 comme référence (un choix est nécessaire de par la non symétrie de la mesure).

La valeur de similarité d'une paire de segments de pulsation est la moyenne de l'addition des similarités entre les onsets appariés, divisée par le nombre d'onsets du segment de pulsation appartenant à la séquence de référence. La valeur de similarité entre deux onsets appariés est la proportion des durées partagées par rapport au nombre total de durées de l'onset appartenant à la séquence de référence. Deux durées $dur1$ et $dur2$ sont similaires, ou partagées,

lorsque :

- les deux durées sont inférieures à 200 ms ;
- les deux durées sont supérieures à 2000 ms et

$$\text{abs}(dur1 - dur2) < 2000 \quad (5.7)$$

- ou bien lorsque

$$\frac{\text{abs}(dur1 - dur2)}{\text{moyenne}(dur1, dur2)} < \text{approx} - \text{dur} \quad (5.8)$$

où ‘approx-dur’ vaut par défaut 1/3.

5.3.7 La similarité globale

Les différentes mesures de similarité que nous venons de décrire sont autant de points de vue différents et autant de valeurs de similarité qu’il faut maintenant combiner pour obtenir, si cela fait sens, une valeur de similarité globale.

Perceptivement, nous pouvons affirmer que la similarité est globale, puisque nous effectuons rarement, ou plus difficilement, des comparaisons selon un seul critère. Par exemple, du strict point de vue du contour, les séquences figure 5.25 sont similaires, mais cela nous est perceptivement difficile à accepter puisque d’autres paramètres (rythme, arrangement polyphonique) nous indiquent le contraire. La perception des contours similaires nécessite un effort particulier qui ne se produit pas forcément pendant l’écoute.

La conséquence est que la qualité de nos différents critères de similarité est difficile à évaluer séparément et requiert une concentration particulière lors des phases de test de l’algorithme. Cependant, la séparation des critères peut aussi être vue comme une richesse puisqu’elle permet de distinguer les rôles respectifs des différents paramètres de similarité qui même perçus globalement ne sont pas forcément explicités par notre perception.

Pour établir une similarité globale, nous avons choisi de combiner linéai-

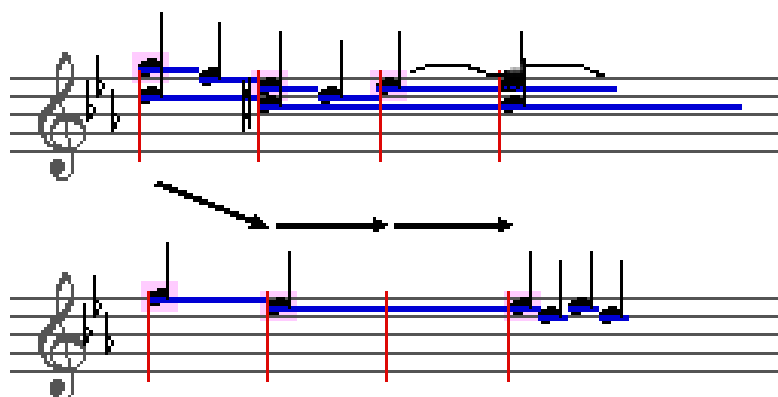


FIG. 5.25: Extraits du *Ricercare à six voix* en Do mineur de Bach. Les contours supérieurs des deux séquences sont identiques, mais les deux séquences sont perçues différentes à cause du contexte rythmique et polyphonique. Un effort particulier est nécessaire pour percevoir les contours identiques.

rement les trois différentes mesures décrites précédemment. Différents poids auraient pu être appliqués à chaque type de similarité, mais puisque nous nous situons dans un contexte le plus général possible, nous avons choisi d'attribuer à chacun des poids la même importance. Un développement possible serait de déterminer une répartition des pondérations optimale issue d'un apprentissage sur un certain nombre de morceaux dont les analyses seraient connues. Une autre idée serait de réaliser une première analyse globale du morceau puis d'utiliser ces résultats pour déterminer lequel de nos trois points de vue serait à privilégier. Par exemple, l'analyse pourrait fournir des indications sur la densité rythmique, la variété des contours ou la densité polyphonique. Un contour peu varié au long de la pièce laisserait supposer que les deux autres aspects sont à privilégier.

5.3.8 La représentation en épures, ou la synthèse par l'analyse

Chacune des mesures de similarité que nous avons défini compare deux séquences suivant un point de vue particulier. Chaque fois, différents éléments des séquences comparées seront mis en relation alors que d'autres ne seront pas considérés pour établir les similarités. Nous proposons alors une représentation des seuls éléments mis en relation, de manière à mesurer l'écart entre les séquences initialement comparées et les éléments de ces séquences ayant servi à établir la similarité.

A chaque type de similarité correspond alors une représentation des éléments mis en relation, que nous appellerons 'épures'. L'épure peut être considérée comme étant le motif lui-même, les notes n'en faisant pas partie n'ayant pas participé à sa détermination. A deux séquences comparées correspondent deux épures, une pour chaque séquence.

Pour la similarité globale, résultant de la combinaison des trois types de similarité (contour, intervalle et rythme), nous pouvons superposer les trois couples d'épures correspondantes en une seule représentation. Les deux épures globales sont alors en quelque sorte un produit de synthèse issu de l'analyse.

Cette représentation est souvent très utile pour l'analyste mais aussi pour le développement de notre algorithme d'analyse car elle nous permet de mesurer la distance qui sépare la musique étudiée du modèle informatique. Un exemple d'épures est donné figure 5.26.

5.4 Récapitulatif

Nous avons défini dans ce chapitre une mesure de similarité applicable à la comparaison de deux séquences polyphoniques ayant même nombre de battues de pulsation. La mesure est constituée de trois valeurs indépendantes caractérisant trois aspects des séquences musicales à comparer : le contour, les intervalles (harmoniques et mélodiques) et le rythme. La combinaison

The image displays musical notation for two sequences and their corresponding epures (skeletons) for comparison. The top row shows 'Sequence 1' and 'Sequence 2' in treble and bass staves. The bottom row shows 'Epure 1' and 'Epure 2' in treble and bass staves. The epures are simplified versions of the sequences, showing only the notes and their positions relative to the pulse lines. The numbers 1, 2, and 10 are placed on the notes to indicate the points of view for comparison: 1 for intervals, 2 for contour, and 10 for rhythm. The epures are aligned with the pulse lines of the sequences above them.

FIG. 5.26: Extraits de la *sonate* en RÉM d664 op121 deuxième mouvement de Schubert. Les deux séquences du haut ont été comparées. Les deux épures issues de la comparaison sont représentées sous les séquences. Chaque épure regroupe les éléments à partir desquels a été établie la similarité. Seuls les éléments situés sur les pulsations sont pris en compte pour le contour et les intervalles. Les numéros sur les épures correspondent aux points de vue : 1 pour les intervalles, 2 pour le contour, et 10 (notes situées en bas des épures) pour le rythme.

de ces aspects donne une valeur globale de similarité. Pour chaque aspect, les évènements ayant contribué à renforcer la valeur de similarité sont retenus et forment ce que nous appelons des épures. Pour chacune des deux séquences comparées, nous disposons ainsi des éléments contribuant à la similarité rythmique, de contour ou d'intervalles. Ces épures nous serviront dans le chapitre suivant à caractériser le 'type' de similarité entre les deux séquences. Le chapitre suivant proposera par ailleurs un outil d'analyse des répétitions contenues dans une séquence entière, que nous représenterons à l'aide d'une matrice de similarité, de laquelle seront extraits les motifs les plus représentatifs. Une solution au problème de la recherche d'un motif donné dans une base de fichiers musicaux sera aussi proposée, ce qui nous amènera à redéfinir notre mesure de similarité pour laquelle la considération des seuls évènements situés sur les pulsations ne sera plus indispensable.

Chapitre 6

L'analyse motivique : l'extraction et la recherche de motifs

Les concepts établis au chapitre précédent nous permettent maintenant de résoudre deux problèmes musicaux bien établis qui sont l'extraction de motifs et la recherche d'un motif donné dans une base de données musicales.

Le premier problème va nous permettre d'introduire une représentation originale des répétitions par le calcul d'une matrice de similarité sur l'ensemble d'une séquence MIDI. Un nouvel algorithme sera alors proposé pour l'extraction de motifs à partir de cette matrice.

Le deuxième problème concerne la recherche de motifs dans une base de donnée MIDI à partir d'un motif pris comme référence. Pour ce problème, nous proposerons une implémentation plus souple de notre mesure de similarité rendue possible par une complexité combinatoire plus faible que celle requise pour le calcul de la matrice de similarité.

Dans l'ensemble du chapitre, nous emploierons le mot 'répétition' pour désigner toutes les séquences perceptivement similaires à d'autres, et le mot 'motif' pour désigner les séquences perceptivement similaires à d'autres mais pouvant être associées à une segmentation de la séquence à analyser, c'est à dire étant perçues comme des entités indépendantes ayant un début et une fin perceptivement marqués.

6.1 Représentation matricielle des similarités

Pour analyser les répétitions contenues dans une séquence musicale dont la pulsation est connue à l'aide de nos trois mesures de similarité, nous avons choisi d'utiliser une matrice de similarité. Pour cela, nous fixons une taille de fenêtre L (en nombre de pulsations), puis nous mesurons dans un ordre chronologique la similarité entre toutes les paires possibles de sous-séquences de taille L contenues dans la séquence à analyser. Nous commençons l'analyse de la séquence trois pulsations avant le premier évènement de manière à se rapprocher de ce que pourrait être notre perception d'un début, caractérisé par le fait qu'aucun évènement ne le précède, ce qui donne une importance particulière aux tout premiers évènements. Chacune des cellules de la matrice contient la valeur de similarité correspondant à la comparaison entre deux sous-séquences de même taille. Les axes de la matrice représentent le temps en nombre de pulsations. Avancer dans la matrice vers le haut ou vers la droite revient donc à avancer dans le temps. Avancer dans une seule direction à la fois (soit horizontalement, soit verticalement) revient à comparer une même séquence de référence avec une séquence que l'on décale d'un segment de pulsation à chaque pas d'avancement. Les coordonnées des cellules correspondent aux positions de fin (en nombre de pulsations) des deux séquences comparées. La concaténation des cellules suivant la diagonale supérieure droite est particulièrement intéressante à observer puisqu'elle correspond à l'agrégation des segments de pulsation consécutifs aux séquences comparées. La taille de la diagonale observée est donc proportionnelle à la taille des séquences comparées. Un exemple d'analyse est donné figure 6.1.

Cette représentation présente l'avantage d'inclure une visualisation des similarités entre séquences qui se recouvrent, ce qui n'est habituellement pas le cas des modèles traditionnels de segmentation tels que ceux par exemple inspirés de Lerddhal et Jackendoff. Nous verrons dans la section intitulée 'Commencer par filtrer les cellules de la matrice' que les recouvrements sont utiles pour extraire des motifs dont la taille dépasse la contenance d'une cellule (soit L pulsations au maximum). Par ailleurs, les recouvrements sont fréquents en

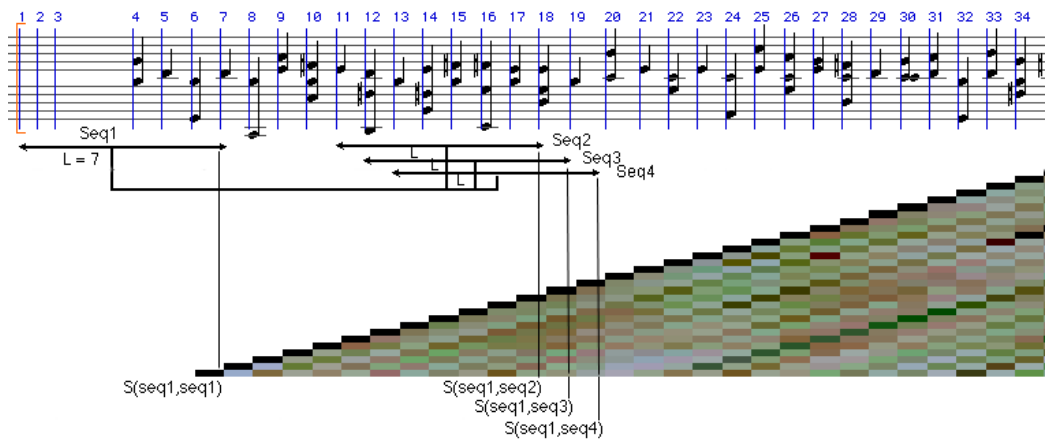


FIG. 6.1: L'analyse du début de la *Pavane pour une infante défunte* de Ravel. Les battues de pulsation sont numérotées et repérées par les barres verticales situées sur les événements de la séquence. Noter que la séquence commence par un silence de trois pulsations. Toutes les sous-séquences de même taille $L = 7$ sont alors comparées entre elles. Les axes de la matrice représentent le temps en nombre de pulsations. Chaque cellule de la matrice correspond à la comparaison entre deux séquences. Les cellules correspondant à la comparaison de Seq1 avec respectivement Seq1, Seq2, Seq3 et Seq4 sont repérées sur la première ligne de la matrice.

musique polyphonique, notamment lorsque le contrepoint est utilisé. Il est donc important de disposer d’une méthode pouvant les mesurer et les représenter.

6.1.1 Pourquoi fixer la taille des séquences à comparer ?

Plutôt que de comparer des séquences de même taille fixe L , nous pourrions choisir de comparer toutes les séquences de toutes les tailles possibles incluses dans le fichier MIDI à analyser. Toutefois, ce serait une opération beaucoup trop lourde en temps de calcul. Par ailleurs, une fois calculées les différentes valeurs de similarité leur grand nombre les rendrait très difficiles à interpréter.

Des séquences de tailles différentes ne seront donc pas comparées, tout comme les séquences dont la taille n’est pas égale à celle que nous aurons choisi. Perceptivement, nous n’opérons sans doute pas de cette façon, puisque la taille des séquences que nous comparons varie plutôt proportionnellement à la taille des motifs déjà mémorisés auxquels on veut les associer.

Il s’agit alors de bien choisir la taille des séquences à comparer, et peut-être d’envisager certaines modifications qui pourraient assouplir cette contrainte. Nous en discuterons dans la section intitulée ‘choisir le bon niveau’.

Pourquoi ne pas mesurer la similarité entre séquences de tailles différentes

Tout d’abord, comme nous l’avons dit, cela induirait un temps de calcul trop coûteux. Par ailleurs, cela poserait un problème très complexe de modélisation. En effet, il faudrait pouvoir associer un groupe de notes à un autre groupe de notes de taille différente et leur trouver une similarité commune en sachant que la solution ne passera pas par une association note à note. Par exemple, la figure 6.2 montre deux séquences similaires dont certains groupes de notes (duolets) devraient être associés à des groupes de taille différente (triolets) si tous les événements des séquences comparées étaient considérés.

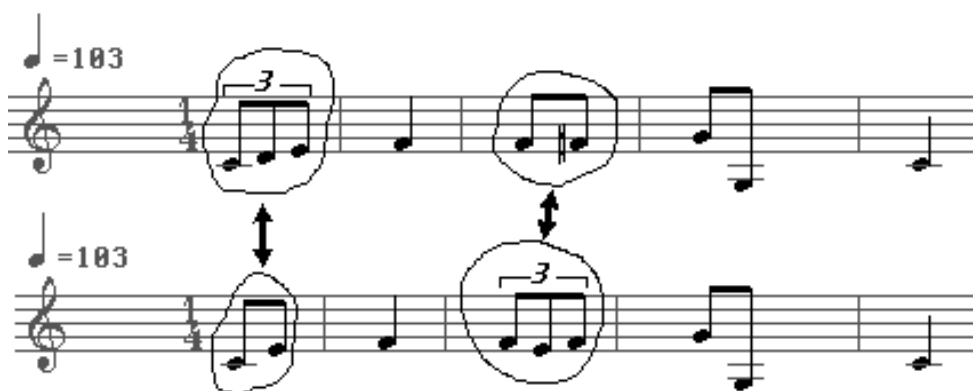


FIG. 6.2: Deux séquences similaires. Certains groupes de notes d'une séquence (duolets) correspondent à des groupes de taille différente (triolets) pour l'autre séquence.

Rolland [44] propose une approche dans ce sens par la programmation dynamique, mais rapidement il doit poser des jalons au système pour ne pas qu'il fasse trop de comparaisons, et ces jalons ne sont pas toujours très convaincants, lorsqu'ils portent par exemple sur la taille maximale des groupes de notes pouvant être comparés, ce qui implique un choix arbitraire. Dans ce cas, une trille comportant une note de plus que la taille autorisée ne serait alors pas considérée. Par ailleurs, les exemples qu'il propose ne prennent pas en compte cette fonctionnalité, ce qui laisse augurer une implémentation difficile.

Une limitation qui autorise certaines variations

La limitation induite par le fait de comparer les seules séquences de même taille est déjà par elle-même réduite puisque la taille n'est calculée qu'en fonction des événements situés sur des battues de pulsation et non en fonction de tous les événements consécutifs comme c'est parfois le cas pour d'autres approches. Les séquences comparées peuvent donc avoir un nombre de notes différent du moment qu'elles ont le même nombre de pulsations. Par exemple,

il est fréquent en musique que des notes de passage ou des ornementsations entre les pulsations soient utilisées pour varier les motifs.

6.1.2 Choisir le bon niveau

Le choix de la taille des séquences à comparer ne doit pas être fait au hasard. Il s'agit de trouver le bon niveau entre la taille maximale possible pour des répétitions qui ne seraient pas reconnues comme telles si elles étaient divisées en plusieurs sous-entités, et la taille minimale à partir de laquelle une séquence peut être considérée comme étant une répétition.

La taille maximale d'une répétition indivisible

Cette notion est assez difficile à définir. Elle s'appuie sur le fait qu'une répétition à notre sens est détectée de manière globale sur les séquences comparées et pas uniquement par concaténation de petites similarités locales à une partie des séquences.

Pour percevoir une répétition, nous faisons intervenir notre mémoire des événements passés de différentes manières : les éléments immédiatement passés, c'est à dire les plus récents sont mis en relation avec les événements actuels pour établir une séquence qui sera comparée à une autre séquence précédemment passée. Ainsi, toutes les relations existantes entre tous les éléments de la séquence actuelle seront comparées avec toutes les relations existantes entre tous les éléments de la séquence passée. Nous avons déjà pris en compte cet aspect en définissant au chapitre précédent des mesures de similarité globales, c'est à dire prenant en compte toutes les relations entre événements à l'intérieur d'une séquence. Nous avons formalisé cet aspect par notre règle de perception 5.1, suivant laquelle la similarité entre deux séquences n'est pas forcément égale à la somme des similarités entre chaque paire d'événements correspondants des deux séquences. Etant donné notre choix de fixer les tailles des séquences à comparer, il nous faut définir une révision de cette règle pour qu'à partir d'une certaine taille, les séquences

puissent être comparées par concaténation. Cette taille, au-delà de sa nécessité algorithmique, peut avoir un sens perceptif. On peut en effet la considérer comme le passage entre un niveau perceptif global et un niveau perceptif local. Ce passage pourrait être lié à notre mémoire à court terme : la taille dépendrait alors de la distance maximale possible entre les deux éléments extrêmes d'une séquence qui interviendraient dans la mesure de similarité et que l'on devrait mémoriser. Au-delà de cette taille, chaque répétition pourrait systématiquement être considérée comme étant la superposition de deux répétitions plus petites (locales) éventuellement se recouvrant. En dessous de cette taille, une répétition ne pourrait pas toujours être divisée sans que son unité soit brisée (on préserve le niveau global), ce dont on se rendrait compte si d'autres répétitions antérieures ne seraient plus considérées comme telles lorsque comparées parties par parties et non plus globalement.

Ce seuil sera matérialisé dans la matrice de similarité par le grain formé par les différentes cellules. On passe d'une cellule correspondant à une similarité globale, à la concaténation d'une suite de cellules différentes (pour que cette concaténation ait un sens par rapport à l'agrégation correspondante des segments de la séquence, il faut suivre la diagonale supérieure droite de la matrice). De cette manière, les variations de similarité entre les différentes parties composantes d'une répétition de taille supérieure à la taille maximale fixée seront visualisables par les variations de couleur des cellules. En dessous de cette taille maximale, seule la similarité globale sera visualisée.

Enfin, précisons que notre concaténation de cellules reste sous l'influence de notre règle de perception 5.1 puisque nous ne concaténons pas des parties disjointes de séquences que nous mettrions bout à bout, mais nous superposons des séquences décalées d'une pulsation ce qui rend les transitions beaucoup plus précises et ce qui assure la prise en compte des relations entre tous les événements de la séquence concaténée sauf celles entre les événements situés aux deux extrêmes. Un exemple est donné figure 6.3.

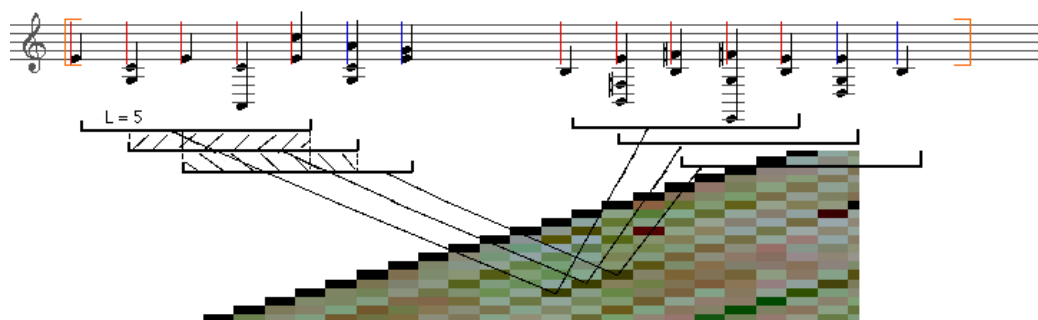


FIG. 6.3: Détail de trois cellules consécutives de la diagonale d'une matrice de similarité. Les trois cellules de la diagonale correspondent à trois comparaisons de paires de séquences de taille 5. Les séquences se recouvrent et sont seulement décalées de un élément à chaque comparaison, si bien qu'une certaine continuité est assurée entre les cellules consécutives de la matrice. De cette manière, les diagonales ne peuvent être vues comme de simples concaténations de cellules consécutives. Sur l'exemple, les parties communes aux différentes sous-séquences de la séquence de gauche sont repérées par des hachures.

La taille comme filtre du nombre de répétitions contenues dans la séquence à analyser

Le choix de la taille des séquences à comparer a aussi une influence sur le nombre de séquences qui seront trouvées similaires.

Si la taille est trop petite, l'algorithme trouvera de trop nombreuses répétitions et beaucoup ne seront pas pertinentes même si perceptivement elles seront jugées similaires.

Inversement, si le nombre de pulsations considéré est trop grand, les répétitions se feront rares, mais elles seront sans doute toutes pertinentes. Le problème sera alors de déterminer combien de répétitions au minimum sont nécessaires pour que l'on juge l'analyse satisfaisante. Ce nombre pourra varier suivant les objectifs de l'analyse. Un exemple de plusieurs analyses faites avec différentes tailles de séquence à comparer est présenté figure 6.4.

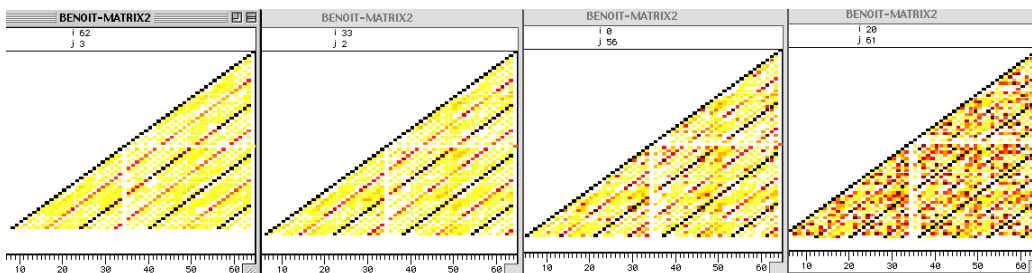


FIG. 6.4: Quatre matrices de similarité d'un même morceau pour quatre tailles de séquence à comparer différentes. De gauche à droite, les tailles sont 9, 7, 5 et 3. Plus la taille diminue, plus les cellules a priori intéressantes, c'est à dire de couleurs foncées, sont nombreuses, mais les diagonales deviennent moins visibles et plus difficiles à interpréter.

Une taille floue

Afin de pouvoir adapter le choix d'une taille à la séquence à analyser, nous avons défini une frontière floue, autorisant 3 tailles possibles en plus de la taille initialement choisie. Les trois nouvelles séquences seront alors obtenues en tronquant progressivement la séquence initiale à partir du début (on conserve les suffixes). Par exemple, en choisissant la taille 7 (en nombre de pulsations), les répétitions de taille 6, 5 et 4 seront aussi considérées. Le fait de conserver les suffixes plutôt que les préfixes peut s'expliquer par le parcours chronologique de la séquence. Chaque déplacement dans la séquence doit correspondre à l'analyse d'un nouvel élément. Or, la taille d'analyse pouvant diminuer, il serait possible en conservant les préfixes de tronquer le nouvel élément à analyser.

A chaque comparaison, la taille qui sera définitivement retenue dépendra des différentes valeurs de similarité trouvées par l'algorithme. Pour avantager les grandes répétitions, souvent plus rares donc plus intéressantes, nous soustrayons à chaque valeur de similarité un poids proportionnel à la taille de la répétition. Par défaut, pour une taille maximale $L = 7$, nous soustrayons les poids 0.1, 0.15 et 0.2 aux similarités obtenues pour les tailles 6, 5 et 4. La prise en compte d'une taille floue fait alors davantage ressortir l'unité

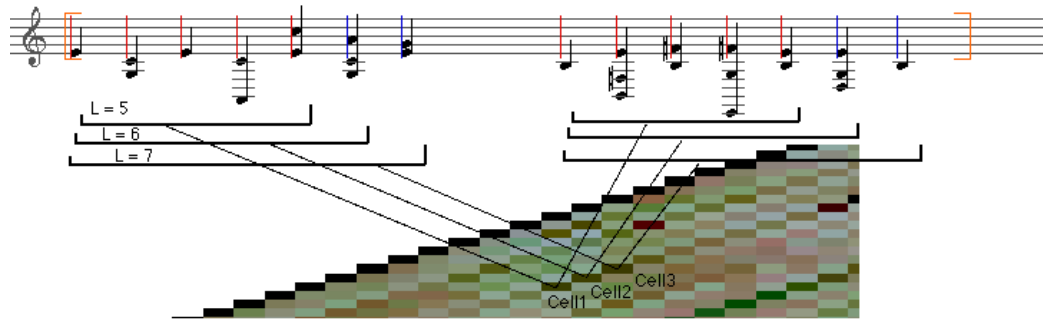


FIG. 6.5: Détail de trois cellules consécutives de la diagonale d'une matrice de similarité, déjà présenté figure 6.3. Les trois cellules de la diagonale correspondent à trois comparaisons de paires de séquences de taille 5 à 7. Le parcours des trois cellules en diagonale montre que les tailles optimales trouvées pour les cellules cell1, cell2 et cell3 sont respectivement 5, 6 et 7. Il y a donc convergence vers une séquence dont le début serait fixe (premier mi de la séquence de gauche et premier si pour la séquence de droite) et la fin non encore déterminée, mais au vu des cellules qui suivent, la similarité (couleur de la cellule) diminue rapidement ce qui signifie que la fin est proche.

des séquences, puisqu'il est possible d'observer la convergence vers une taille fixe de séquence lors du parcours d'une diagonale de la matrice de similarité. L'exemple donné précédemment figure 6.3 est repris figure 6.5 en considérant cette fois une taille floue pouvant aller de 5 à 7.

Dans les parties suivantes, nous supposons qu'une taille L (en nombre de pulsations) a été choisie. Toutes les séquences qui seront comparées auront alors une même taille comprise entre $L-3$ et L .

6.1.3 Le calcul d'une matrice - synthèse

Notre matrice de similarité contient Nb lignes et Nb colonnes. Nb est le nombre total de pulsations contenues dans la séquence à analyser. Chaque cellule contient le résultat de la comparaison entre deux séquences de taille L , qui est composé de : une valeur de similarité, la taille des séquences comparées correspondantes (identique pour les deux séquences mais pouvant varier

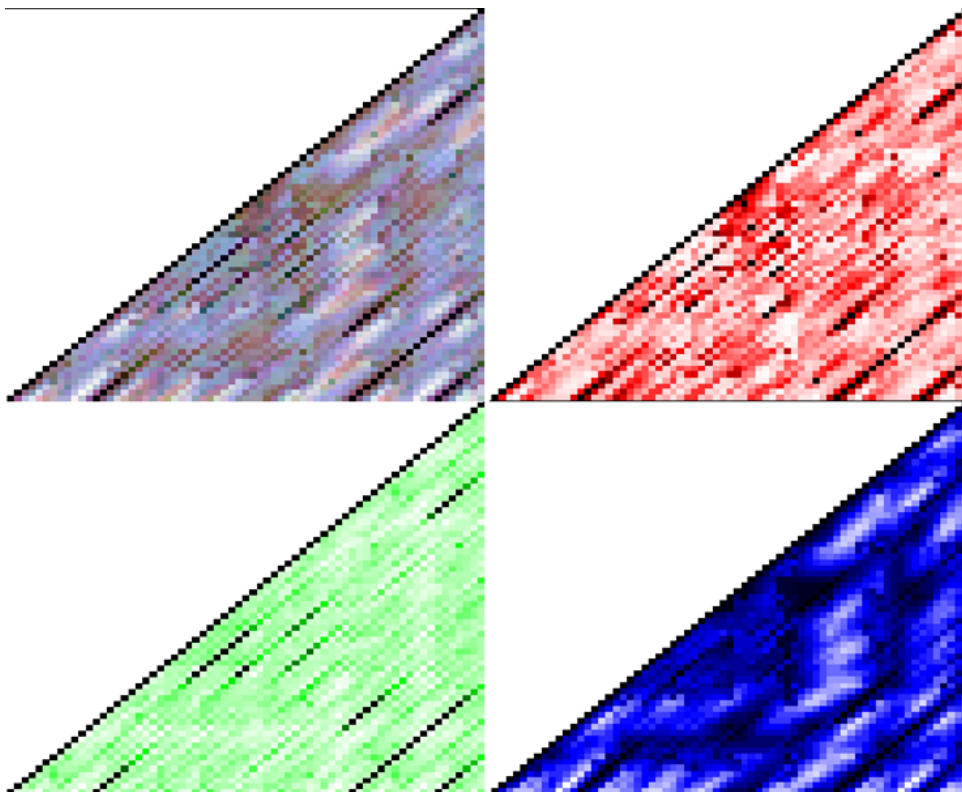


FIG. 6.6: Les différentes matrices extraites de la Sonate en Ré Majeur d664 op121 3ème mouvement de Schubert. Les matrices nous renseignent chacune sur un aspect du morceau.

de L-3 à L), ainsi que les épures issues de la comparaison (nous avons défini la notion d'épure au chapitre précédent intitulé 'la synthèse par l'analyse'). Les diagonales supérieures droites sont particulièrement intéressantes puisqu'elles correspondent à la comparaison entre deux séquences contenues dans le morceau à analyser de taille supérieure à celle définie pour l'analyse.

Pour une même analyse, il est possible de visualiser les différentes matrices correspondant aux différentes similarités : contour, intervalles, rythme, similarité globale. Parfois, il est alors possible de mettre en évidence de grandes différences ou de grandes similitudes dans l'emploi de ces différents points de vue sur l'ensemble du morceau. Par exemple, la figure 6.6 montre les quatre différentes matrices (contour, intervalle, rythme et similarité globale)

obtenues à partir de l'analyse de la Sonate en Ré Majeur d664 op121 3ème mouvement de Schubert. La matrice de similarité d'intervalles présente les plus forts contrastes. En effet, cette mesure considère la similarité entre lignes horizontales et verticales d'intervalles identiques à une transposition près, ce qui fait que les répétitions totalement exactes sont presque les seules à être détectées. La matrice d'intervalle représente alors surtout la forme du morceau en fonction de ses reprises, mais parfois des répétitions internes à la polyphonie ou propre aux enchaînements d'accords peuvent aussi être détectées. La matrice de contour est difficile à interpréter pour notre exemple, mais elle présente souvent une sorte d'affinement de la matrice d'intervalle par la prise en compte des répétitions non exactes. L'association matrice d'intervalle et matrice de contour permet alors de conserver une hiérarchie de structure entre les grandes lignes de reprise du morceau et les similarités plus locales. La matrice de rythme montre des transitions rythmiques assez progressives qui provoquent une impression de représentation en trois dimensions. Cela provient du fait que des rythmes similaires décalés de quelques unités peuvent rester similaires, ou deviennent différents progressivement, alors qu'un profil mélodique $m1$ similaire à un profil $m2$ peut être complètement différent de ' $m2$ décalé d'une pulsation'. Les rythmes sont souvent composés de la répétition d'une même petite cellule ce qui est moins vrai pour les mélodies, comme l'illustre la figure 6.7.

Par ailleurs, nous avons choisi pour nos expériences de ne pas représenter la non symétrie de la mesure de similarité (définie par l'équation 5.3) pour faciliter l'interprétation des résultats, même si cet aspect a été pris en compte par nos mesures de similarité : nous ne retenons que la plus forte valeur des deux mesures, et les matrices de similarité deviennent alors triangulaires.

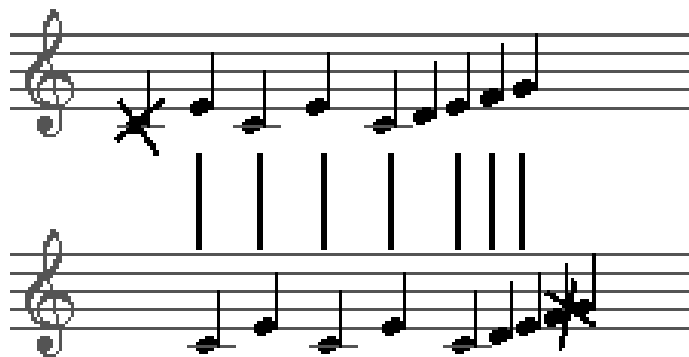


FIG. 6.7: Deux séquences identiques que l'on compare après avoir décalé la deuxième d'une pulsation. Les séquences sont plus ou moins différentes suivant que l'on considère le rythme ou la mélodie : les rythmes restent très proches et chaque nouveau décalage les éloignerait de façon continue. Les contours par contre deviennent opposés dès le premier décalage. La transition dans la similarité se fait donc de façon discontinue, sauf lorsqu'on suit une diagonale supérieure droite qui correspond à un décalage d'une pulsation pour chacune des deux séquences. Ces différences de comportement apparaissent sur nos différentes matrices de similarité.

6.2 L'extraction de motifs à partir de la matrice de similarité

Une fois la matrice calculée, différents types d'informations peuvent en être extraits, par exemple l'évolution générale du rythme, des intervalles de hauteurs ou du contour, la répétition locale de cellules ou les points de rupture dans la structure. Il est possible à chaque fois de relier ces informations avec la notion de motif. Cependant, nous allons dans un premier temps nous intéresser à la seule extraction de ce qu'on peut appeler les motifs 'les plus importants'.

6.2.1 Quels sont les motifs 'les plus importants'

Les motifs 'les plus importants' sont assez difficiles à définir, puisqu'il n'existe pas de critère objectif permettant de les caractériser. Nous allons donc considérer à nouveau le point de vue de la perception.

On pourrait par exemple penser que perceptivement, puisque les motifs se caractérisent par leurs répétitions, le terme 'les plus importants' désigne les motifs qui sont répétés dans la séquence musicale le plus grand nombre de fois. Seulement, du fait même qu'il est répété très souvent, un motif perd de son importance perceptive. Par exemple, le même accompagnement peut se retrouver tout au long d'une séquence musicale. Pour autant, ce motif n'apparaîtra pas toujours comme étant perceptivement le plus important puisque en tant qu'accompagnement régulier il se situe en second plan. Par ailleurs, les motifs les plus souvent répétés sont souvent les plus petits (dans la limite de taille que nous nous sommes fixés), alors que nous aurions une tendance à considérer les plus grands comme plus importants. Mais de même, rechercher les motifs les plus grands ne nous amène pas forcément aux 'plus importants', puisque de par leur taille, ils ne sont pas mémorisés aussi facilement que les motifs de taille 'moyenne', qui leur sont donc préférés.

Peut-être faut-il rechercher successivement les motifs suivant ces différents critères, puis les organiser selon le rôle qu'on leur attribue.

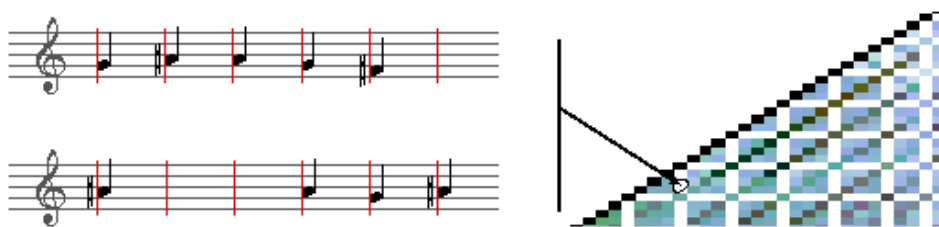


FIG. 6.8: Deux séquences qui ne seront pas comparées, car l'une se termine par un silence et l'autre pas. Cela revient donc à analyser les deux séquences tronquées de leur dernier élément. Or, cette comparaison a déjà été effectuée lors d'une étape précédente. Le résultat est inscrit dans la cellule voisine inférieure gauche. Pour éviter les redondances dans la matrice, nous laissons alors la cellule blanche.

6.2.2 Commencer par filtrer les cellules de la matrice

La lecture des motifs à partir de la matrice n'est pas immédiate. En effet, chaque cellule correspond à la comparaison entre deux séquences dont la taille peut varier entre $L-3$ et L . Les diagonales supérieures droites formées par une suite de cellules représentent des comparaisons de séquences de taille au moins supérieure à $L-2$, la plus grande diagonale représentant la séquence musicale entière comparée à elle-même.

La taille réelle d'une séquence musicale en nombre de pulsations correspondant à une suite de N cellules est donc variable. Elle peut aller de $N + L - 4$ à $N + L - 1$.

Par ailleurs, lorsque deux séquences vont être comparées, et que l'une d'entre elles se termine par un silence, la comparaison n'est pas effectuée, car elle serait redondante avec la comparaison des deux séquences dont la plus longue aurait été tronquée de son dernier élément. Sur la matrice, nous représentons ce cas de figure pas une case blanche, signifiant que les séquences n'ont pas été comparées. Par contre, deux séquences se terminant par un silence seront comparées, car l'algorithme peut éventuellement aboutir à un résultat différent de celui de la comparaison entre les deux séquences ne contenant pas les silences. Un exemple est donné figure 6.8.

La lecture d'une diagonale doit quand même pouvoir se faire dans le cas où par exemple deux motifs seraient similaires, mais l'un contiendrait un silence et pas l'autre. Cela apparaîtrait sur la matrice par une diagonale comportant une case blanche sur la position du silence. Cette case ne devra alors pas être interprétée comme une frontière ou une rupture dans le motif, mais plutôt comme un élément de variation inclus dans le motif.

Le cas des silences traité, chaque cellule de la matrice correspond alors à une comparaison unique entre deux séquences, puisque à deux fins données de séquences à comparer il correspondra une unique cellule de la matrice.

Enfin, du fait que chaque cellule de la matrice est associée à des séquences de plus d'une pulsation, des cellules consécutives correspondent toujours à des séquences qui se chevauchent, voir qui se recouvrent. Cette représentation, assez riche mais complexe, est difficile à analyser lorsqu'on cherche à en extraire des motifs dont la taille dépasse la contenance d'une cellule, soit L pulsations au maximum. Il faut en effet alors concaténer les cellules voisines formant une diagonale supérieure droite. Ces cellules peuvent avoir une couleur différente et il s'agit alors de déterminer si les couleurs sont suffisamment continues pour pouvoir les associer entre elles.

Préalablement à l'extraction de motifs, nous définissons un seuil de similarité en dessous duquel les cellules ne seront plus prises en compte ce qui permet de faire davantage ressortir les diagonales de cellules correspondant à des séquences fortement similaires. Par défaut, nous avons fixé ce seuil tel que $1/10$ des cellules de la matrice soit conservé. Un exemple est montré figure 6.9.

6.2.3 Etablir une classification des cellules d'une même ligne

Nous supposons maintenant que la matrice de similarité ne contient que les cellules dont la valeur de similarité est supérieure à un certain seuil fixé. Nous proposons dans cette section un algorithme de classification des cellules d'une même ligne. En effet, chaque ligne de la matrice correspond à la com-

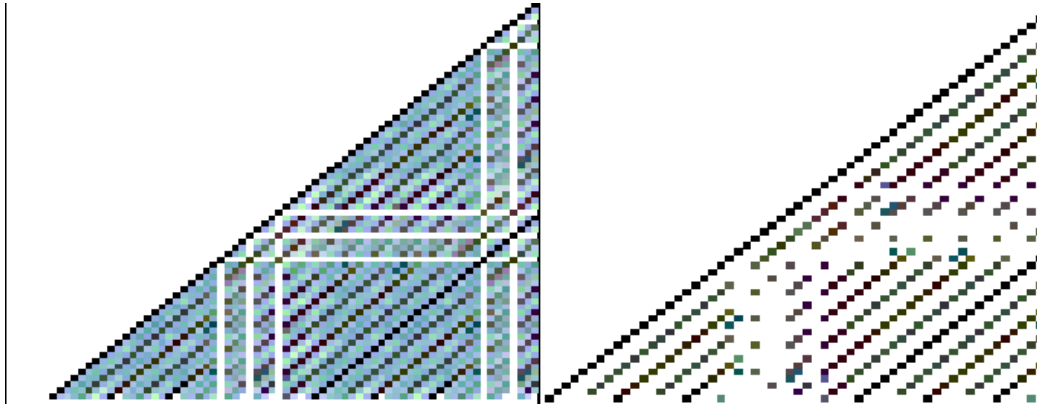


FIG. 6.9: Deux matrices de similarité. Celle de gauche n'a pas été filtrée. Seul 1/10 des cellules ont été conservées dans celle de droite. Les diagonales apparaissent plus nettement et sont plus aisément reconnaissables par l'informatique.

paraison d'une séquence de référence avec toutes les séquences de même taille qui la suivent dans le temps. Chaque cellule de la ligne représente le résultat d'une comparaison. On veut alors classer les différentes cellules selon la similarité qu'elles entretiennent avec la séquence de référence. Il s'agit d'établir un type de similarité, en quelque sorte une similarité sur les similarités. Pour cela, nous allons nous baser sur les comparaisons entre les épures associées aux cellules à classer.

En effet, à chaque mesure de similarité entre deux séquences correspondent deux épures que l'on peut interpréter comme représentantes du type de la similarité puisqu'elles contiennent les notes ayant servi à caractériser la similarité entre les deux séquences. Par exemple, à une séquence donnée $S1$ comparée avec une séquence $S2$ correspondront deux épures $E(S1,1)$ et $E(S2,1)$. A la même séquence $S1$ comparée avec une autre séquence $S3$ correspondront deux autres épures $E(S1,2)$ et $E(S3,1)$. $E(S1,1)$ et $E(S1,2)$ seront deux épures différentes mais toutes deux associées à $S1$ et donc composées de certains évènements de $S1$ (voir figure 6.10). En supposant ces deux épures similaires, on pourrait déduire que $S1$ entretient avec $S2$ et $S3$ une similarité de même type. Donc, les deux épures de $S1$ pourraient être regroupées dans

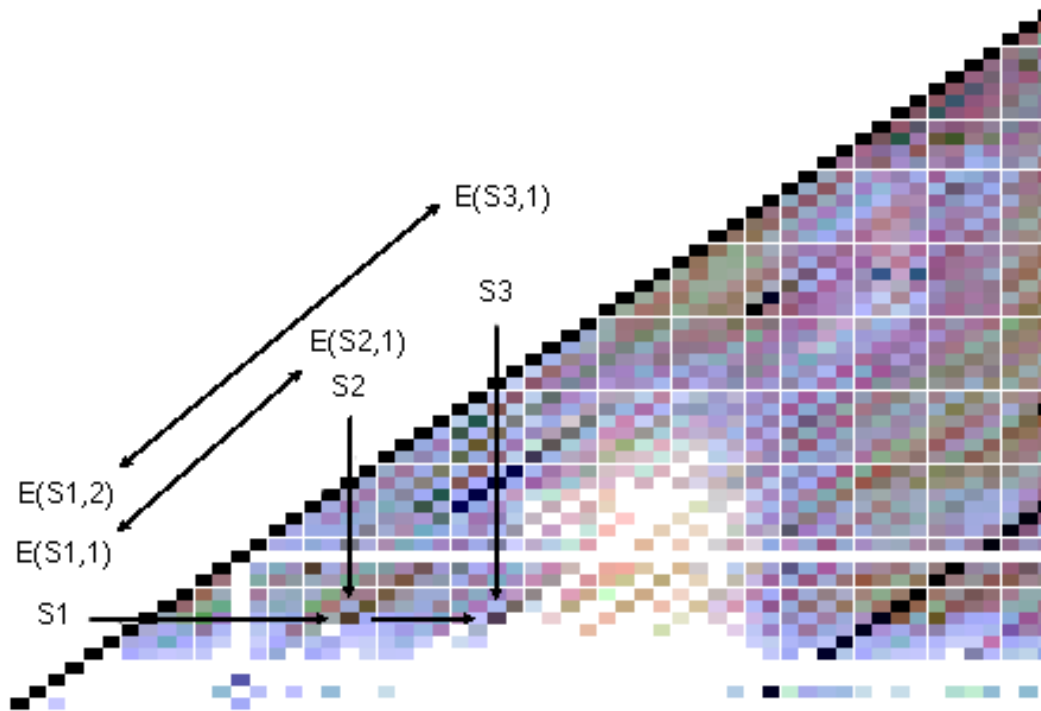


FIG. 6.10: Analyse d'un extrait d'un prélude de Bach. Représentation des épures issues de la comparaison entre la séquence S1 (diagonale débutant sur la cellule noire marquée S1) et les séquences S2 et S3 (diagonales débutant à l'intersection des lignes verticales S2, S3 avec la ligne horizontale S1 de la matrice). Les différentes épures ($E(S1,2)$ et $E(S1,1)$) associées à S1 peuvent être classées par type. Pour cela, nous évaluons leur similarité.

un même ensemble. Il reste maintenant à adopter une mesure de similarité entre épures et à définir notre algorithme de classification.

Nous proposons d'utiliser les mêmes mesures de similarité que celles que nous avons employé pour les séquences, et d'adopter une valeur à partir de laquelle les deux épures seront dites similaires. Cette valeur doit être assez grande pour être sélective car les épures comparées étant issues de la même séquence, elles sont toutes plus ou moins similaires. Par défaut, nous avons fixé cette valeur à 0.7. Pour simplifier la classification, et étant donné que nous nous situons au niveau abstrait de l'épure, nous supposons que la relation de transitivité est vraie. Donc, si deux épures E1 et E2 sont similaires (de même

type) et si E2 est aussi de même type que E3, alors E1 et E3 seront considérées de même type. Pour chaque ligne, nous obtenons alors différents groupements d'épures. Nous sommes maintenant prêts à commencer l'extraction de motifs proprement dite.

6.2.4 Comment extraire les motifs du plus petit au plus grand

Nous proposons maintenant d'extraire les suites de cellules de la matrice qui correspondent à des motifs.

Nous partons des classes d'épures de la première ligne de la matrice que nous cherchons à propager vers les lignes supérieures. Prenons l'exemple de la figure 6.11. A chaque ligne 'correspond' un motif de référence, représenté par la cellule noire à l'extrémité gauche de la figure, la ligne étant le résultat de la comparaison du motif de référence avec tous les motifs qui lui sont consécutifs.

La ligne horizontale du bas de la matrice contient huit cellules. Seules 6 de ces cellules ont été regroupées ensemble du fait de leur 'même' similarité avec la cellule de référence (en noir à gauche). Si nous montons maintenant d'un cran (deuxième ligne horizontale), nous pouvons concaténer les 6 cellules avec leurs 6 cellules consécutives (en suivant la diagonale supérieure droite). Nous avons donc extrait 6 motifs de taille 2 cellules similaires au motif de référence de taille 2 contenu dans la diagonale extrême supérieure gauche de la matrice. Comme la troisième ligne ne compte pas de nouvelle cellule, le processus s'arrête à ce stade. Si une cellule avait prolongé l'un des six motifs, nous aurions eu un nouveau motif de taille trois associé au prolongement du motif de référence contenu dans la diagonale extrême gauche. Ainsi de suite, nous aurions construit un arbre dont le nombre de branches pour chaque nouvelle ligne explorée n'aurait pu que rester stable ou diminuer. A partir de l'élément de référence 'correspondant' à la première ligne, nous construisons ainsi un réseau de motifs, du plus petit au plus grand, le nombre de motifs ayant tendance à diminuer à chaque nouveau pas (à chaque

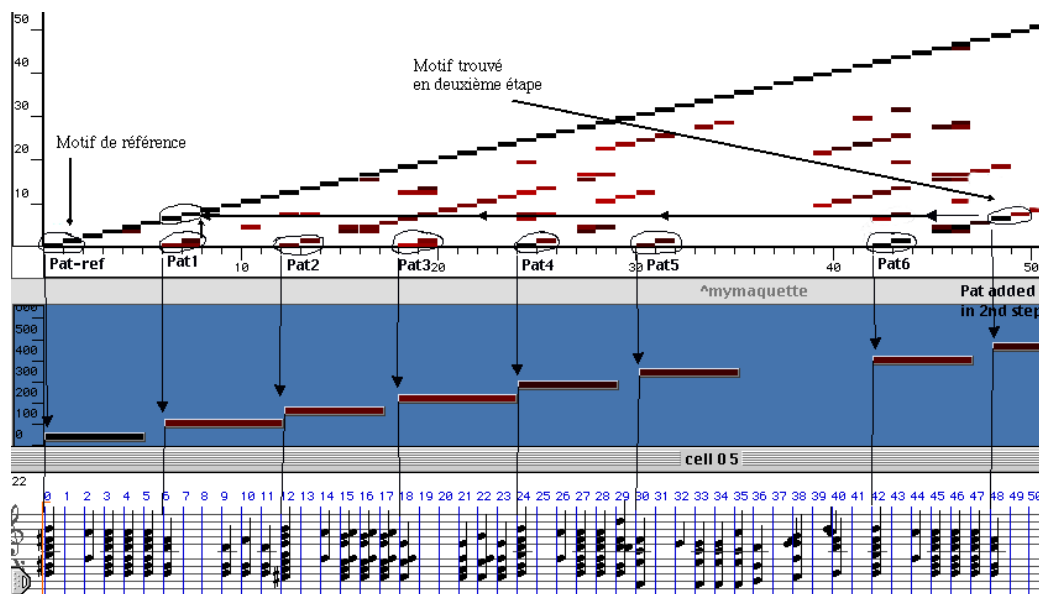


FIG. 6.11: Successivement de haut en bas : la matrice de similarité calculée à partir de la Sonate en Ré Majeur d664 op120 2ème mouvement de Schubert, les motifs qui ont été extraits (chaque segment horizontal est une instance du même motif) et la séquence musicale correspondante (les axes de la matrice et la grille de la séquence musicale sont exprimés en ‘nombre de segments de pulsation’).

nouvelle ligne explorée). Revenons maintenant à notre exemple. Nous avons trouvé six motifs similaires au motif de référence de taille 2 commençant à la première ligne. Dans un deuxième temps, nous associons alors à chacun des six motifs la ligne horizontale qui lui ‘correspond’. Ces lignes contiennent des groupes de cellules qui peuvent à leur tour être comparées à notre motif de référence. En cas de ‘même’ similarité, ces cellules sont rajoutées à notre réseau. Figure 6.11, nous avons surligné les 6 motifs extraits dans un premier temps, ainsi qu’un motif trouvé dans le deuxième temps de l’extraction (le dernier sur la droite de la figure) à partir de la ligne horizontale du premier des six motifs.

Une fois le réseau de la première ligne constitué, nous construisons un nouveau réseau en prenant cette fois pour référence la deuxième ligne de la matrice, et ainsi de suite. Les réseaux sont ensuite comparés. Certains étant inclus dans d’autres leurs sont alors agrégés.

Au final, nous obtenons un ensemble de réseaux, identifiés par un motif de référence (le plus grand du réseau) et un ensemble de sous-motifs du motif de référence, chaque motif étant associé à plusieurs occurrences similaires.

6.2.5 Comment extraire les motifs tout court

Pour extraire les motifs les plus pertinents, nous les classons par taille. En plus de cette contrainte, nous sélectionnons les réseaux dont le motif de référence ne recouvre pas le motif qui lui est associé de plus de ‘taux-recouvr’ exprimé en pourcentage de la taille du motif. Une représentation graphique du réseau dont le motif de référence est le plus grand est alors proposée par l’algorithme.

Figure 6.12, nous avons représenté un réseau constitué de trois niveaux. Le réseau peut être vu comme une segmentation de la séquence analysée. L’axe horizontal de la figure correspond à l’axe du temps. Le premier niveau (en bas de la figure) correspond au motif de référence (en noir à gauche) associé à un motif qui lui est similaire. Le deuxième niveau correspond à un sous-motif du motif de référence (en noir au-dessus du motif de référence) associé à 4

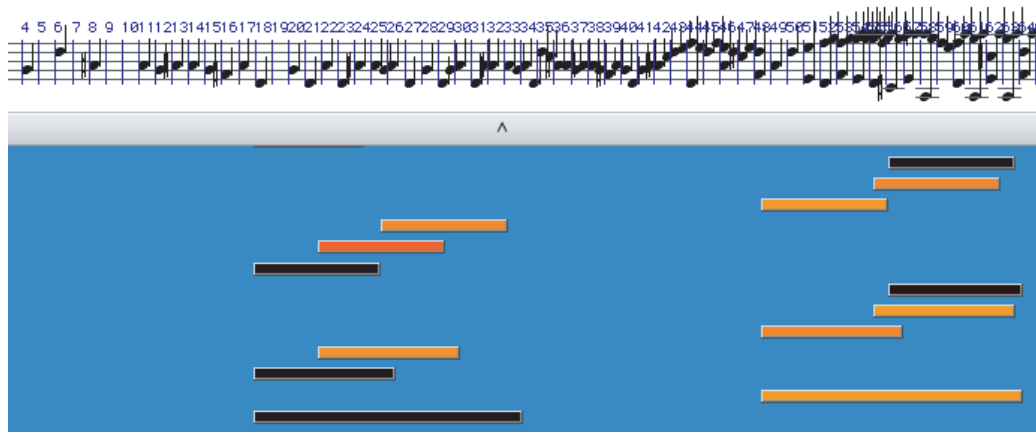


FIG. 6.12: Un réseau de motifs positionné sous les éléments de la séquence musicale lui correspondant. Le réseau constitue une possible segmentation de la séquence. L'élément de référence du réseau, en noir au bas à gauche, est similaire à l'élément supérieur décalé à droite. En allant vers le haut, on raffine progressivement le motif de référence par des éléments plus petits mais plus nombreux.

motifs qui lui sont postérieurs. Nous obtenons ainsi un raffinement progressif du découpage, partant d'unités de grande taille mais peu nombreuses, et allant vers des unités de plus petite taille mais plus nombreuses.

6.2.6 Définir les débuts et fins de motifs

Nous n'avons pas encore abordé le problème de déterminer les débuts et fins de motif. En effet, les séquences extraites de la matrice ne sont pas forcément perçues comme une totalité par rapport à la séquence entière, et donc ne sont pas forcément des motifs, même si de par leur construction elles sont plus proches de la notion de motif qu'une simple répétition prise au hasard dans une cellule de la matrice.

Le problème de début et fin consiste à définir le seuil de similarité à partir duquel on peut considérer une ligne de la matrice comme étant une position de début ou de fin de motif. Nous avons déjà vu dans la section consacrée aux 'tailles floues' des séquences à comparer que certains types de

répétitions permettent d'induire le début d'un motif. En effet, lorsque pour plusieurs cellules consécutives d'une diagonale d'une matrice de similarité, la meilleure taille de séquence croît d'une unité à chaque pas, nous pouvons supposer que le début correspond au début commun à l'ensemble de ces séquences comparées (voir la figure 6.5).

Par ailleurs, si nous avons considéré que les accents n'étaient pas très utiles pour extraire des répétitions, il pourraient être employés pour déterminer à l'intérieur de ces répétitions certains événements plus importants que d'autres provoquant une segmentation pouvant correspondre à un début ou une fin de motif. Cependant, nous n'avons pas défini plus précisément quels marquages pourraient être utilisés, et nous considérons donc toutes les diagonales de la matrice filtrée comme étant des motifs possibles, les extrémités des diagonales constituant les débuts et les fins des motifs les plus longs.

6.3 L'extraction de motifs à partir d'une séquence de référence

Cette section propose d'appliquer nos mesures de similarité à l'étude des variations d'un motif donné à l'intérieur d'un corpus de fichiers MIDI.

6.3.1 Une complexité moindre

Ce problème est moins complexe que celui de l'extraction de motifs à partir d'une matrice de similarité puisqu'il s'agit maintenant de comparer une seule séquence, préalablement donnée, avec les séquences potentiellement similaires contenues dans un ou plusieurs fichiers MIDI. Pour une séquence donnée de taille L et un fichier MIDI de N événements, l'algorithme effectue donc $4 * (N+1-L)$ comparaisons en supposant que seules les séquences de même taille, en nombre de pulsations, soient comparées.

La complexité étant très faible, nous pouvons donc affiner notre recherche de similarité, en particulier en autorisant les séquences comparées à avoir des

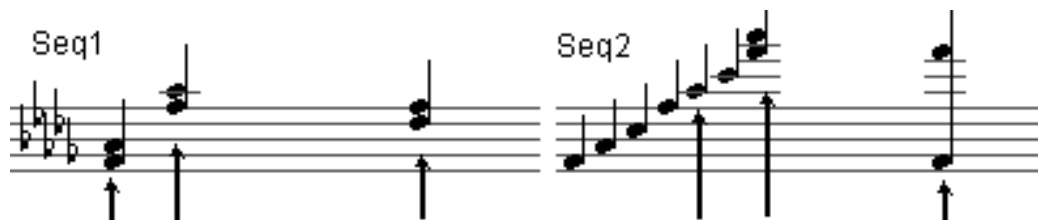


FIG. 6.13: Deux séquences extraites du *Clair de lune* de Debussy. Les deux séquences repérées par des flèches n'ont pas un rythme identique, et sont situées dans un contexte polyphonique différent mais elles sont pourtant détectées comme similaires par notre nouvel algorithme ce qui semble perceptivement satisfaisant.

tailles différentes. Nous ne nous baserons alors plus sur les pulsations, mais sur le nombre d'évènements pour mesurer la taille des séquences. Cette modification apporte beaucoup de souplesse à l'algorithme car il n'est plus nécessaire de connaître préalablement une pulsation. Par ailleurs, si les séquences une fois mises en relation ont toujours la même taille (chaque évènement d'une séquence correspond à un unique évènement de l'autre séquence), elle peuvent être issues de séquences de tailles très différentes, dont les évènements ne sont pas nécessairement situés sur les pulsations, comme le montre la figure 6.13.

Pour un fichier MIDI donné, la comparaison entre une séquence $S1$ et toutes les séquences possibles contenues dans le fichier ne ferait pas vraiment sens puisque des séquences du fichier dont les évènements sont très distants les uns des autres pourraient être formées. Nous définissons donc une taille limite L à ne pas dépasser, puis nous déplaçons une fenêtre de taille L dans le fichier MIDI évènement par évènement. A chaque pas, la séquence la plus similaire à $S1$ contenue dans la fenêtre est retenue. Pour éviter les redondances, le dernier évènement de la fenêtre est toujours considéré dans la comparaison. La comparaison entre la séquence $S1$ et une séquence de même nombre d'évènements se fait en utilisant les mesures de similarité définies au chapitre précédent.

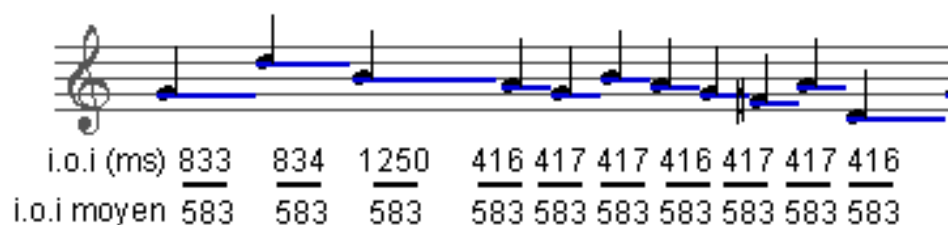


FIG. 6.14: Un profil rythmique est composé de la liste des inter-onsets divisés par l'inter-onset moyen de la séquence.

La comparaison entre une séquence S_1 comprenant n_1 événements et les différentes séquences inclues dans un fichier MIDI de taille N et en considérant une fenêtre de taille $L \geq n_1$ demandera alors $(N-L) * C(n_1) (L-1)$ mesures de similarité entre deux séquences de même nombre d'évènements.

6.3.2 Préserver une cohérence rythmique

Pour limiter le temps de calcul et pour préserver une certaine cohérence rythmique, nous imposons comme contrainte que les séquences comparées aient des rythmes similaires d'après la définition suivante :

Nous définissons deux types de rythmes similaires :

- toute séquence dont les éléments sont consécutifs est considérée comme ayant un rythme 'similaire' à la séquence de référence ;
- les autres séquences ont un rythme similaire à la séquence de référence si leurs profils rythmiques lui sont proportionnels.

Le premier cas est immédiat à traiter. Pour le second, nous allons définir ce qu'est un profil rythmique.

Pour une séquence MIDI donnée, nous définissons le profil rythmique comme étant la suite des inter-onsets de la séquence divisés par l'inter-onset moyen calculé sur l'ensemble de la séquence. Un exemple est donné figure 6.14.

Deux profils rythmiques sont similaires si la différence entre chaque élément se correspondant n'excède pas $2/9$. Par ailleurs, N variations de plus de $2/9$ mais de moins de $3/9$ sont autorisées. N est un nombre fixé en fonction de la taille n en nombre d'évènements des séquences comparées. Par défaut, N est initialisé à $(n - 1) / 2$. Par cette représentation, des rythmes de tempo différent pourront être jugés similaires sans la connaissance du tempo.

Si deux profils rythmiques sont reconnus similaires, la similarité globale entre les séquences sera évaluée, mais une pénalité rythmique pondérera le résultat final suivant la valeur de similarité entre les profils rythmiques :

$$penalite = 1 - \frac{\frac{1}{4} \times (\text{nombre de notes différentes})}{N} \quad (6.1)$$

D'autres formes de profils rythmiques pourraient être adoptées, par exemple en utilisant 'l'inter-onset interval ratio', qui consiste à diviser chaque inter-onset par l'inter-onset précédent. Notre représentation a cependant l'avantage d'exprimer les inter-onsets en fonction de la taille globale de la séquence ce qui permet de rapprocher des séquences de proportions globalement similaires même si des différences peuvent être envisagées localement.

6.3.3 Un outil pour établir des liens entre différents morceaux

L'application de cette mesure de similarité à un ensemble de fichiers MIDI permet d'établir des liens inédits entre différentes séquences. Le motif polyphonique de référence peut être considéré comme point de départ puis en fonction des résultats, il est possible de naviguer dans les différents fichiers en choisissant à chaque fois un nouveau motif de référence parmi ceux trouvés. On peut aussi conserver le motif de référence et établir un classement des fichiers MIDI selon les occurrences trouvées du motif. Un exemple est donné figure 6.15. Le motif de référence est extrait du deuxième mouvement de la 'sonate n°3 op.2' de Beethoven. Les premiers et troisième motifs de la figure sont intéressants mais le deuxième l'est beaucoup moins, ce qui montre

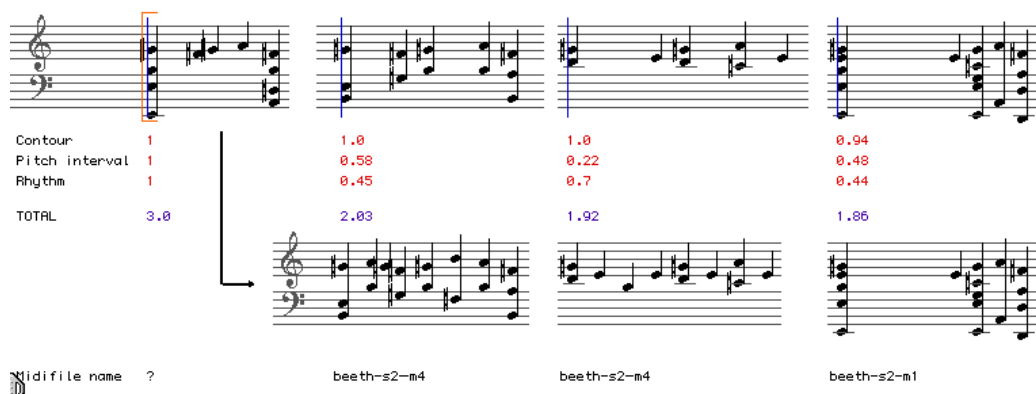


FIG. 6.15: Résultat de la recherche automatique d'un motif dans un corpus de sonates de Beethoven. Le motif de référence, à gauche de la figure, est tiré de la 'sonate n°3 op.2 2ème mvt' de Beethoven. Des motifs similaires ont été trouvés dans les premiers et quatrièmes mouvements de la 'sonate n°2 op.2' de Beethoven. Les motifs sont classés par similarité décroissante de gauche à droite, avec pour chaque motif, en haut de la figure les épures (notes ayant servi à établir la similarité) et en bas toutes les notes du motif.

la difficulté d'extraire d'une polyphonie des motifs réellement perçus comme tels. Le troisième motif est extrait du Quatrième mouvement de la 'sonate n°2 op.2' de Beethoven. Ce motif est important puisqu'il est développé tout au long du mouvement ce qui permet d'établir un lien très intéressant entre ce mouvement et le mouvement du motif de référence.

Il serait possible d'affiner la représentation des résultats en proposant le choix du classement des motifs similaires suivant les trois différents points de vue de nos mesures de similarité. La suite des variations du motif initial serait alors davantage contrôlée, ce qui constituerait un pas vers une application compositionnelle.

6.4 Récapitulatif

Ce chapitre et le précédent ont été consacrés à l’analyse motivique. A partir d’une séquence symbolique polyphonique ou monophonique dont la pulsation est connue, nous calculons une matrice de similarité dans laquelle nous pouvons visualiser différents types de répétitions. A partir de cette matrice, certains réseaux de motifs considérés comme les plus pertinents par l’algorithme sont extraits automatiquement, et représentés sous la forme de segments superposés à la séquence musicale. La lecture de bas en haut des segments permet d’avancer dans le niveau de définition, les segments du bas correspondants aux motifs les plus longs mais souvent les moins répétés, alors que les segments du haut, sous-parties des segments les plus longs, sont plus petits mais plus souvent répétés.

Nous proposons par ailleurs une nouvelle définition de notre mesure de similarité pour effectuer la recherche d’un motif donné dans une base de données. Le motif peut être choisi par l’utilisateur, ou être issu de l’un des réseaux de motifs extrait précédemment par l’algorithme de la matrice de similarité. Dans sa nouvelle définition, la similarité s’établit non plus pour les seuls événements situés sur les battues de pulsation, mais pour tous les événements de la séquence. Ceci est rendu possible par la plus faible complexité combinatoire du nouveau problème. L’outil informatique permet alors d’effectuer des rapprochements intéressants entre différentes œuvres, quelles soient du même auteur ou d’auteurs différents. C’est un pas vers la caractérisation du style d’un auteur, rendu explicite par la recherche d’un vocabulaire de motifs qui lui sont spécifiques et qui se trouvent employés dans plusieurs de ses œuvres.

A partir de nos algorithmes d’analyse, nous avons étudié 500 fichiers MIDI dont plusieurs présentaient des variations de tempo. L’analyse motivique s’est faite sur les 70 premières pulsations de chaque fichier. Le temps total d’analyse automatique (pulsations et motifs) a pris 10 heures, soit un peu plus d’une minute par morceau. L’analyse de la pulsation présente un temps négligeable par rapport à celle des motifs. Les résultats nous sont parus

très pertinents à l'écoute. Nous allons maintenant dans le chapitre suivant confronter nos algorithmes à deux exemples musicaux issus de différents styles du répertoire classique.

Chapitre 7

Synthèse et exemples

Dans ce chapitre, nous présentons les différents modules d'analyse décrits précédemment à travers deux exemples musicaux. Tout d'abord, les principales fonctions de chaque module sont résumées brièvement et les différents paramètres utilisés pour nos analyses sont précisés. Ensuite, deux analyses sont présentées, puis nous nous appuyons sur les différents résultats pour effectuer une première évaluation critique de notre système.

7.1 Le modèle général d'analyse

Une analyse consiste à extraire d'un fichier MIDI dont sont connus les onsets, hauteurs, et durées, une représentation du fichier par matrice de similarité à laquelle sont associés un ou plusieurs motifs.

L'analyse se fait en quatre temps :

- formatage du fichier MIDI ;
- recherche d'une pulsation et d'une métrique ;
- calcul de la matrice de similarité ;
- extraction de motifs représentatifs et recherche de liens avec d'autres morceaux.

Toutes ces fonctions d'analyse ont été implémentées dans Open-Music, un environnement de composition assistée par ordinateur, principalement

destiné aux compositeurs [1]. Il est écrit en CLOS (Common Lisp Object System, extension objet de CommonLisp) qui est un puissant langage à la fois fonctionnel et objet.

7.1.1 Formatage du fichier MIDI

Les fichiers MIDI donnés en entrée de l'analyse peuvent être assez différents suivant leurs origines. Il est donc nécessaire de passer par une étape de formatage pendant laquelle on s'assure que les différents paramètres nécessaires à l'analyse sont présents et que d'autres paramètres dont nous n'avons pas besoin n'interféreront pas avec les fonctions de l'analyse. Par ailleurs, les fichiers, s'ils proviennent d'une interprétation, peuvent contenir des irrégularités rythmiques, trop faibles pour être associées à des variations de tempo, mais susceptibles d'induire nos algorithmes en erreur.

Vérification des paramètres

Pour chaque fichier MIDI donné en entrée, nous ne retenons que les informations sur les onsets, durées et hauteurs. Par exemple, dans l'hypothèse où différents canaux auraient été établis, nous les regroupons tous en un seul. Certains fichiers peuvent contenir des informations de type percussion contenues dans le canal 10. Ces informations n'ont pas le même statut que les autres données MIDI, puisqu'elles ne sont associées à aucune hauteur et durée réelles. Par ailleurs, leur donnée résoudrait dans la plupart des cas de manière triviale notre problème d'extraction de la pulsation. Nous les éliminons donc des fichiers lorsque le cas se présente.

Correction des faibles erreurs temporelles

Lorsque les fichiers sont issus d'une interprétation, les notes des accords peuvent être légèrement décalées dans le temps. Dans ce cas, l'accord sera considéré par l'algorithme comme une succession de différentes notes ce qui pourrait fausser les résultats. Nous opérons donc une première quantification

temporelle de façon à ce que les notes séparées de quelques millisecondes soient rassemblées en un seul évènement. Par défaut, nous regroupons les notes lorsqu'elles sont distantes de moins de 50 millisecondes. Certains morceaux seront alors parfois déformés, par exemple en cas de trilles ou d'ornements très rapides, mais détecter ce cas de figure demanderait une analyse trop complexe pour un simple formatage de données, aussi nous laisserons ce problème de côté. Enfin, seules les 50 premières secondes de chaque morceau seront considérées dans nos trois analyses.

7.1.2 Recherche d'une pulsation

Nous appliquons ensuite au fichier MIDI notre algorithme d'extraction de pulsation. Pour chacune des différentes étapes les valeurs par défaut des fonctions d'analyse sont fixées à l'avance et ne seront pas variées suivant les morceaux à analyser.

Valeurs par défaut des marquages (f1)

Le marquage₁ attribue le poids 1 ou 0. Deux inter-onsets sont similaires si leur différence n'excède pas 0.2 fois leur moyenne.

Le marquage₂ attribue un poids variant de 0 si l'inter-onset courant n'excède pas 1.2 fois l'inter-onset le précédant à 10 si l'inter-onset courant excède de 1.2 fois les 10 inter-onset consécutifs le précédant.

Le marquage₃ attribue un poids de 0 aux silences ou aux accords formés d'une note, 2 aux accords de 2 notes, et 4 aux autres accords

Le marquage₄ attribue un poids variant de 0 si l'onset courant ne correspond à aucune fin de durée hormis celle de l'onset le précédant, jusqu'à 10 s'il correspond à la fin de durée de 10 ou plus onsets le précédant. L'onset o_2 correspond à la fin de la durée d_1 dont l'origine est l'onset o_1 si : d_1 ne diffère pas de $o_2 - o_1$ plus de 0.1 fois $o_2 - o_1$.

Les marquages ne sont ni normés ni pondérés avant leur utilisation pour l'extraction de pulsation.

Induction de pulsation (f2)

La taille ‘taille-contexte’ de la séquence à partir de laquelle de fait l’induction de pulsation est fixée à 3000 ms. Le nombre maximal de pulsations autorisées ‘nb-puls-max’ est fixé à 9. La période maximale ‘pulse-max’ pour une pulsation est de 2100 ms.

Propagation de pulsation (f3)

La taille de fenêtre ‘taille-approx’ dans laquelle les battues de pulsations peuvent correspondre approximativement à la battue attendue est fixée à 0.2 fois la période attendue.

Mise à jour de la l.p.p

‘Contexte-pulse’, le nombre maximal de battues consécutives qu’une pulsation peut avoir sans être mise à jour, est fixé à 8.

Extraction de la meilleure pulsation (f4)

Nous avons retenu les critères 1 et 3 pour classer les différentes pulsations :

- le critère 1 consiste pour chaque pulsation à diviser le nombre de battues trouvées par la moyenne de la période, puis à classer les pulsations selon les valeurs obtenues dans l’ordre décroissant ;
- le critère 4 consiste pour chaque pulsation à rechercher les autres pulsations qui lui sont multiples. Deux pulsations sont multiples entre elles si leurs périodes ont un rapport de proportion simple et si leurs battues sont en phase. Les rapports de proportion retenus sont $1/2$, $1/3$, 2, 3 et 4. Deux périodes sont en rapport de proportion si lorsque l’une est ramenée à la même proportion que l’autre, les deux périodes ne diffèrent pas de plus d’une proportion de $1/10$. Des battues sont en phase lorsque elles coïncident sur au moins $2/3$ du nombre de battues de la pulsation en ayant le moins. Nous classons ensuite les pulsations par ordre décroissant du nombre de multiples qu’elles possèdent. En cas

d'égalité, le classement précédent sera inchangé et interviendra donc dans le résultat final.

Nous retenons alors la meilleure pulsation du classement sauf dans un cas de figure précis : lorsque le tempo attribué à la meilleure pulsation dépasse 300 et que la deuxième pulsation du classement est dans un rapport de proportion $1/2$ ou $1/3$ par rapport à elle, la deuxième pulsation sera retenue. Au final, nous obtenons une séquence dont chaque battue de pulsation est connue. La séquence est régularisée de manière à ce que toutes les battues soient situées à intervalles réguliers.

7.1.3 Extraction de la métrique

Les marquages

Pour chacun des cinq marquages (dynamique, ambitus, durée, nombre de notes), mis à part M3 (silence), un poids normé entre 0 et 8 est donné proportionnellement aux différentes valeurs des propriétés correspondantes. Le marquage M3 qui est booléen est pondéré par 0 ou 8.

Le marquage dynamique prend la plus grande valeur de dynamique (allant de 0 à 127) pour un évènement donné.

Le marquage ambitus prend la différence entre les hauteurs extrêmes d'un évènement exprimées en midi cents.

Le marquage silence (booléen) marque les évènements dont toutes les durées sont inférieures de 50 millisecondes à leur inter-onset avec l'évènement suivant.

Le marquage durée prend la plus grande durée (en millisecondes) pour un évènement donné.

Le marquage densité prend le nombre de notes de l'évènement courant.

La fonction d'autocorrélation

La fonction d'autocorrélation prend une taille de fenêtre de 'p' * 'lag-presume' où 'p' est le nombre de périodes à considérer et 'lag-presume' est

la période supposée par défaut (en nombre de pulsations). Par défaut, ‘p’ = 7 et ‘lag-presume’ = 1. Une fois le graphe d’autocorrélation établi, on le parcourt chronologiquement à la recherche des pics localement maximaux. Un pic situé à la position ‘pos’ du graphe est localement maximal si :

- tous les pics qui le précèdent lui sont inférieurs ;
- les pics qui le suivent dans un intervalle de $1/3$ de ‘pos’ lui sont inférieurs ;
- les pics qui le suivent dans un intervalle de ‘pos’ ne lui sont pas supérieurs de plus de $2/3$.

Les positions des pics retenus dans le graphe d’autocorrélation déterminent les périodes possibles de mesure. Pour chaque calcul de graphe d’autocorrélation, plusieurs périodes sont donc envisagées. La période la plus souvent envisagée pour l’ensemble des graphes est finalement retenue.

7.1.4 Mesure des similarités

Nous avons déjà détaillé précédemment les valeurs employées pour les trois mesures de similarité correspondant aux trois points de vue adoptés : contour, intervalles mélodiques et harmoniques, et rythme. Pour le rythme, le paramètre ‘approx-on’ est fixé à 0.1. Il correspond à la différence tolérée entre deux positions d’onsets rapportées à la taille du segment de pulsation dans lequel elles sont contenues.

La similarité globale est calculée en donnant la même importance à chaque point de vue. Elle est donc calculée en faisant une simple combinaison linéaire des trois mesures de similarité.

7.1.5 Calcul de la matrice de similarité

Nous fixons la taille de la séquence à analyser à 70 pulsations. Les séquences comparées ont une taille de 7 pulsations. Soit $S(\text{seq1}, \text{seq2}, L)$ la similarité entre deux séquences à comparer de taille L . La taille de seq1 et seq2 est définie par :

$\text{taille}(\text{seq1}, \text{seq2}) = \text{taille-seq-max} ((S(\text{seq1}, \text{seq2}, 7)), (S(\text{seq1}, \text{seq2}, 6) - 0.1), (S(\text{seq1}, \text{seq2}, 5) - 0.15), (S(\text{seq1}, \text{seq2}, 4) - 0.2))$

où taille-seq-max désigne la taille associée à la valeur de similarité maximale

7.1.6 Extraction de motifs représentatifs et recherche de liens

Le filtrage de la matrice est fait en ne conservant que 1/10 des cellules (les cellules aux plus fortes valeurs de similarité sont gardées). Pour les groupements de cellules d'une ligne horizontale ayant même 'type' de similarité, nous posons le seuil de similarité entre les épures à 0.7. L'extraction des meilleurs réseaux de motifs contenus dans les diagonales de la matrice se fait suivant trois critères : la taille des plus grands motifs du réseau, le degré de recouvrement entre les deux plus grand motifs du réseau (les motifs ne doivent pas se recouvrir de plus de 1/3) et la taille du réseau en nombre de motifs.

7.2 Analyse du début de la 'Sonate n°9' de Beethoven

7.2.1 Recherche d'une pulsation

Le fichier utilisé provient d'un enregistrement MIDI contenant des variations de tempo. Sans étape de quantification préalable, les courbes de pulsation trouvées seraient assez irrégulières et les rapports harmoniques entre les pulsations seraient difficiles à établir. Après quantification préalable à l'extraction du rythme, les courbes de pulsation sont plus régulières. Le fichier est alors donné en entrée de l'algorithme d'extraction de pulsation. La figure 7.1 montre les deux résultats obtenus avant et après quantification préalable.

Les courbes de pulsation sont plus cohérentes (plus régulières et en rap-

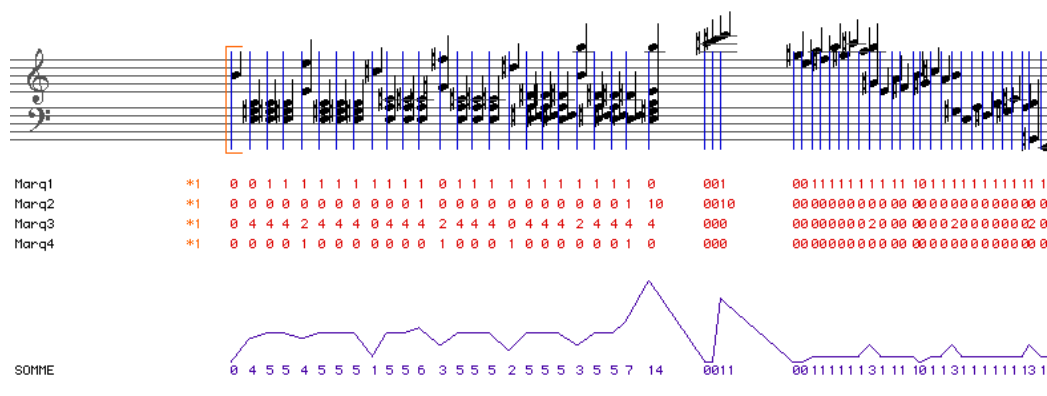


FIG. 7.2: Les quatre marquages utilisés pour l'extraction de pulsation.

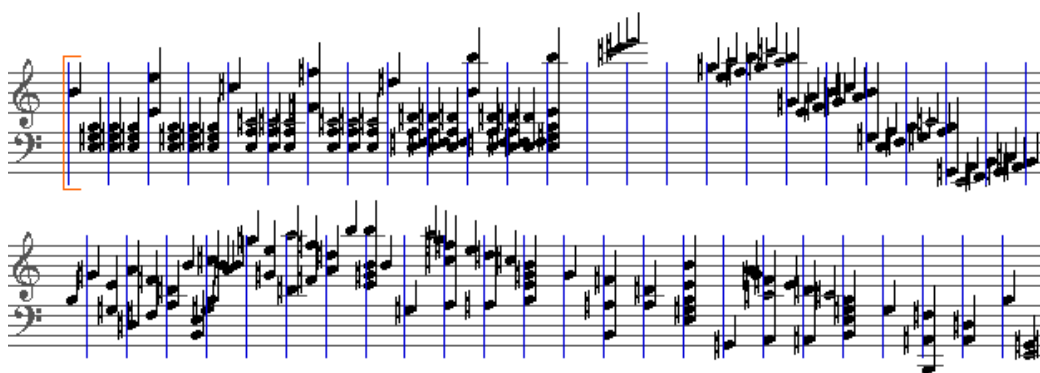


FIG. 7.3: La séquence quantifiée après extraction de la pulsation. Chaque barre verticale coïncide avec une battue de pulsation.

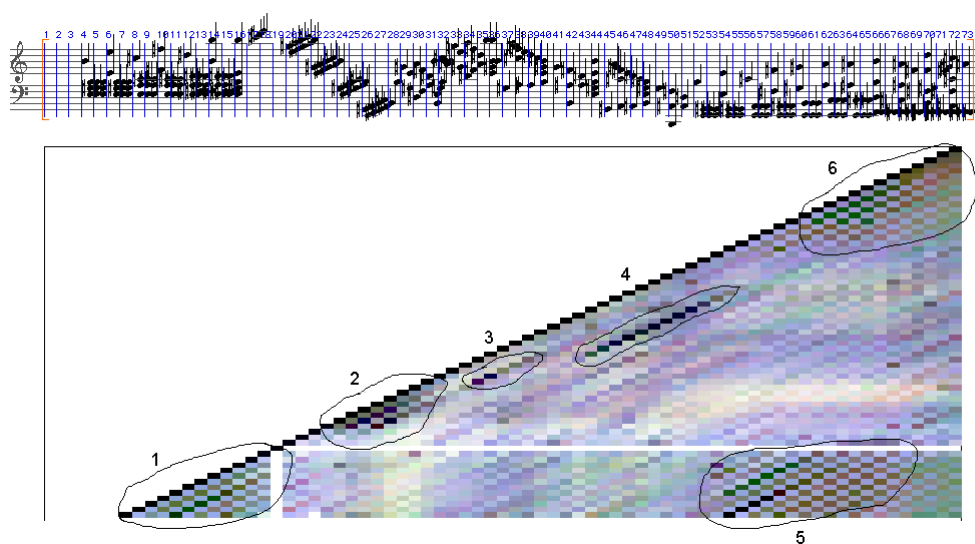


FIG. 7.5: La matrice de similarité calculée sur les 70 premières pulsations.

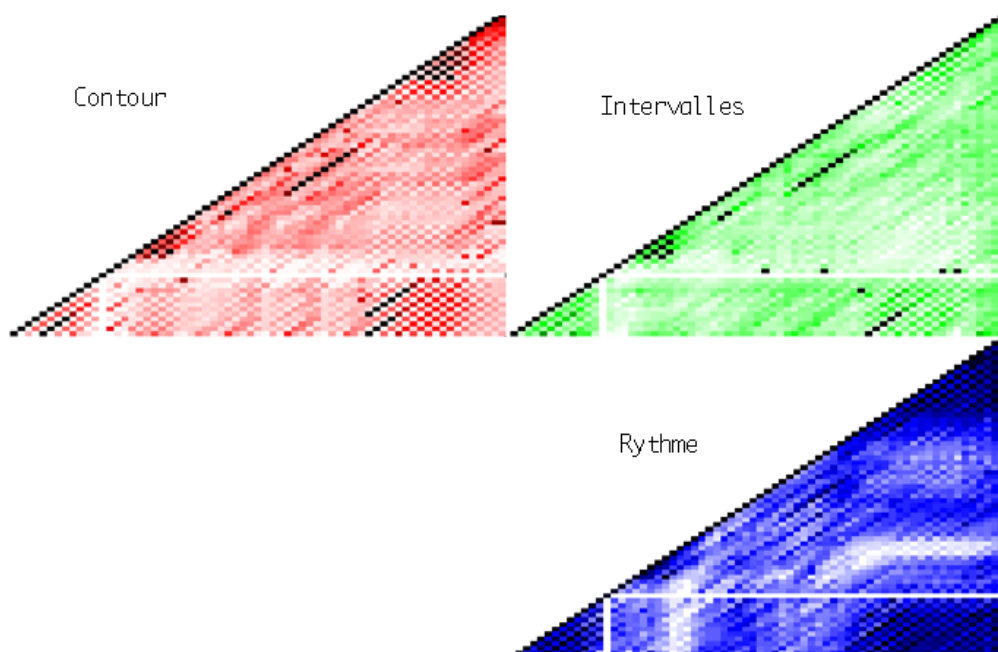


FIG. 7.6: Les trois matrices correspondant respectivement à la similarité par contour, intervalles et rythme.

foncée de la matrice numérotée 1, jusqu'au trait blanc vertical), puis un bref instant il n'y a plus de répétitions jusqu'à ce que les petits groupements de notes soient repris dans une descente régulière (numérotée 2 vers la 20ème barre de pulsation), ce qui se voit dans la matrice par des diagonales sombres proches de la diagonale délimitant l'extrémité supérieure de la matrice. On sait que ces événements sont sans rapports avec le début, car les premières lignes de la matrice restent très claires. Vient ensuite une nouvelle répétition (en mouvement ascendant) d'une séquence plus longue, répétée une fois (numérotée 3, une seule diagonale sombre dans la matrice), puis encore une nouvelle formule plus longue répétée en descendant (numérotée 4, une seule diagonale sombre plus longue que les précédentes). Le thème initial est alors repris, ce qui se traduit dans la matrice par des diagonales sombres dans les premières lignes (numéroté 5). Ce thème lui-même contient des répétitions, ce qui fait qu'une zone triangulaire sombre apparaît en haut de la matrice (numéroté 6).

Le détail des différentes valeurs obtenues pour la similarité par contour, intervalle et rythme est représenté figure 7.6. Les influences respectives du rythme, du contour et des intervalles apparaissent alors assez nettement au regard de l'organisation générale du morceau que nous venons de décrire.

7.2.4 Extraction motivique

Les cellules de la matrice sont filtrées d'après leurs valeurs de similarité de manière à ne conserver qu'une proportion de 1/10 d'entre elles correspondant aux cellules ayant la plus forte valeur de similarité. La matrice obtenue est alors représentée figure 7.7. Certaines diagonales ressortent, mais d'autres restent difficiles à interpréter ce qui met en relief la difficulté d'interprétation automatique de la matrice. Les différentes parties que nous avons décrites à partir de la matrice générale apparaissent cependant assez nettement et révèlent avec plus de contraste la structure du morceau.

Le résultat de l'algorithme d'extraction motivique de cette matrice est représenté figure 7.8. De bas en haut sont représentés les motifs du réseau jugé

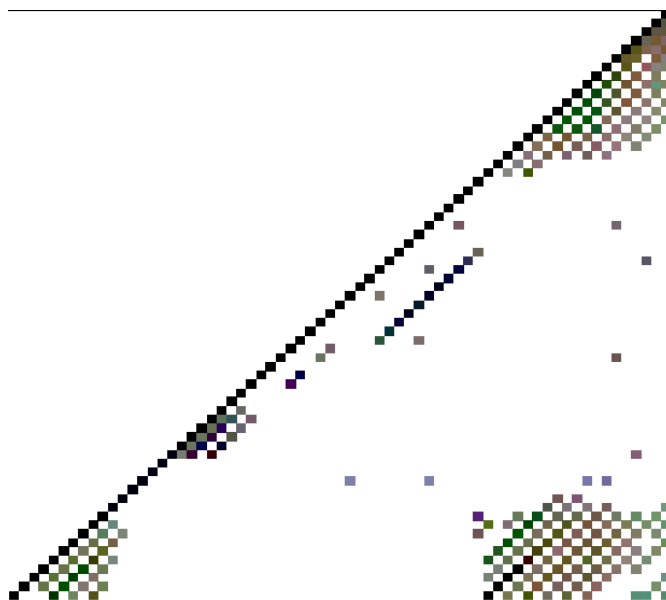


FIG. 7.7: La matrice de similarité filtrée. Seulement 1/10 des cellules les plus similaires ont été conservées.

le plus pertinent par l'algorithme, du plus grand motif (le motif de référence) au plus petit. Chaque motif de couleur noire à gauche est similaire à une succession de motifs qui lui sont postérieurs dans la séquence. Ce résultat revient sur la matrice de similarité figure 7.5 à associer le début de la diagonale principale (dans la zone bas gauche de la matrice générale numérotée 1) à la plus grande diagonale située en bas à droite de la matrice (numérotée 5). Une autre diagonale est intéressante (milieu haut de la matrice, numérotée 4), mais est constituée de la répétition d'un motif plus petit, donc jugé moins intéressant par l'algorithme.

7.3 Le début de la 'Pavane pour une infante défunte' de Ravel

Nous allons maintenant montrer un exemple de recherche d'un motif donné dans une base de données. Nous allons partir de la *Pavane pour une*

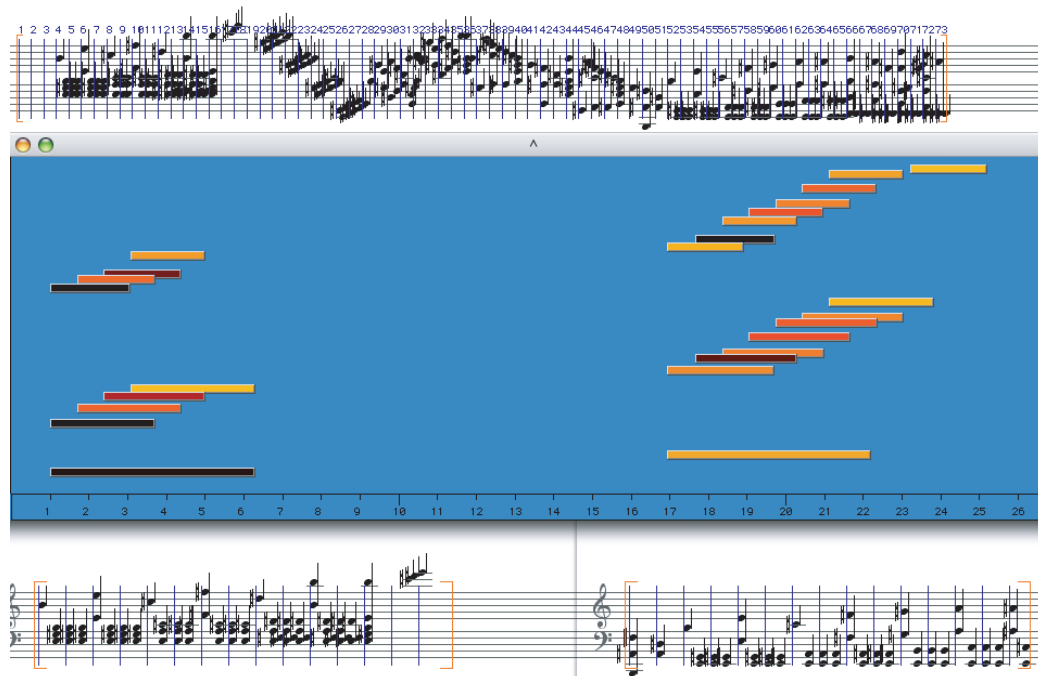


FIG. 7.8: Le réseau de motifs jugé le plus pertinent par l'algorithme. Il correspond à un découpage cohérent de la partition (première phrase montante jusqu'à la descente par petits groupes de notes, reprise à la fin de la séquence). Les deux motifs principaux (bas de la figure) sont transposés et diffèrent sur la fin, mais ils restent perceptivement similaires.

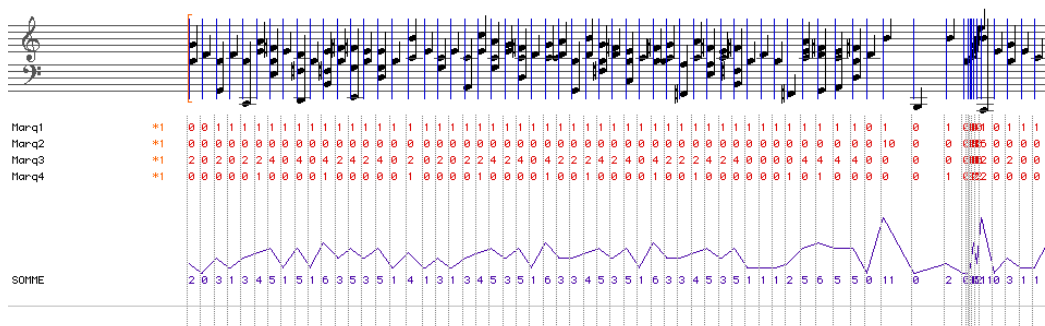


FIG. 7.9: Marquage utilisé pour l'extraction de pulsation.

infante défunte de Ravel pour laquelle nous présentons d'abord brièvement l'analyse rythmique et la matrice de similarité calculée par l'algorithme.

7.3.1 Recherche d'une pulsation

Le marquage du début de la séquence est représenté figure 7.9.

La l.p.p (ou courbe des pulsations) est représentée figure 7.10. Le tempo retenu est de 107 à la noire. La courbe des différentes pulsations montre bien la cohérence entre les différentes valeurs proposées par l'algorithme puisque les irrégularités d'une pulsation se retrouvent de façon presque identique chez les autres pulsations commençant au début de la séquence. Les faibles périodes qui apparaissent vers la fin (à droite dans le graphe) sont induites par la passage correspondant aux notes rapides ascendantes. Ce passage est très court, ce qui fait que sur le graphe, les périodes induites ne seront pas propagées longtemps et disparaîtront rapidement.

7.3.2 Calcul de la matrice de similarité

La matrice de similarité est représentée figure 7.11. On observe une coupure nette vers la fin de la séquence (partie verticale de cellules claires, nu-

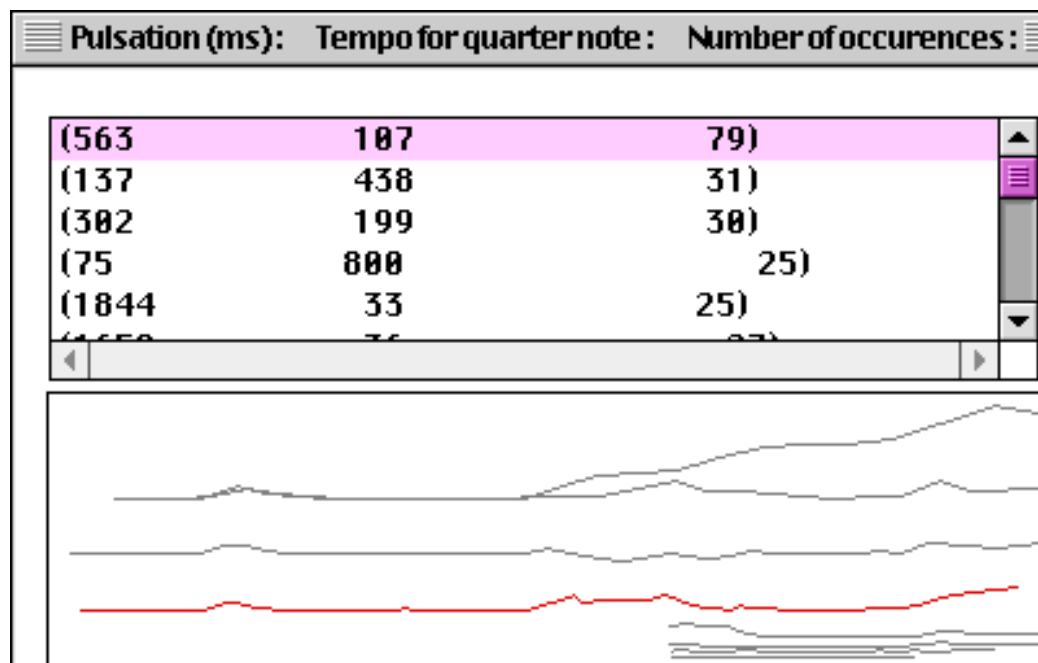


FIG. 7.10: Courbe de tempo des différentes pulsations proposées par l'algorithme.

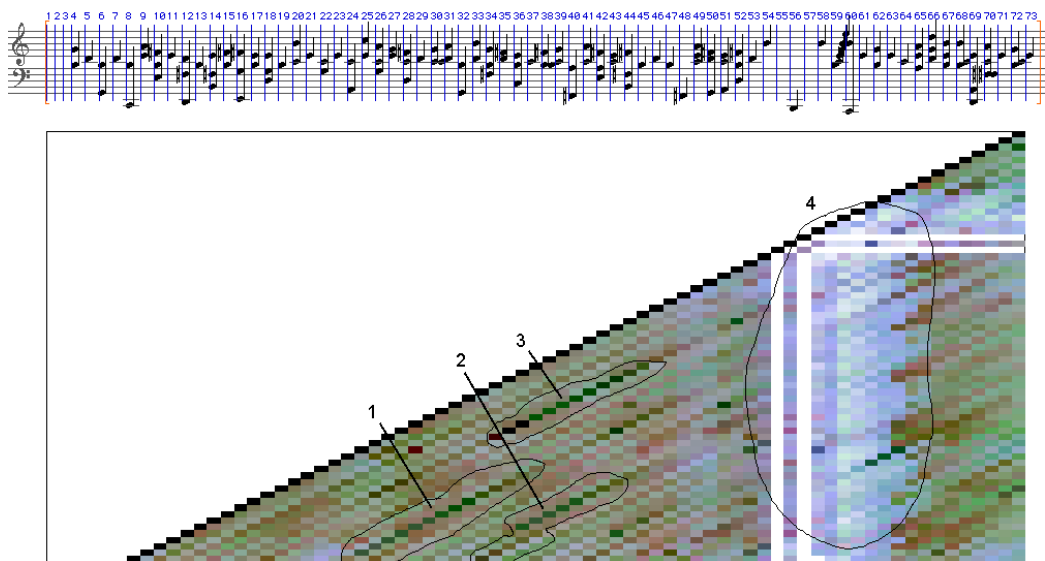


FIG. 7.11: Matrice de similarité calculée sur les 70 premières pulsations.

mérotée 4), qui correspond bien à un contraste perceptible : la régularité du morceau (régularité rythmique, de contour et intervalique) est brusquement interrompue, puis reprend progressivement. Les diagonales assez foncées (numérotées 1, 2 et 3) correspondent au thème initial, qui est repris deux fois. La diagonale 1 est alors la première répétition du thème, la diagonale 3 est la première répétition de la reprise du thème, alors que la diagonale 2 représente la deuxième répétition du thème. Cette dernière répétition comporte sur sa fin un décalage rythmique qui donne l'impression d'un déséquilibre, ce qui se traduit par une diagonale en zig-zag (en fait, la fin du thème a été tronquée, enchaînant avec une nouvelle répétition, comme si le temps s'était accéléré de façon discontinue).

7.3.3 Extraction motivique

Le réseau de motifs retenu par l'algorithme est représenté figure 7.12.

Nous sélectionnons alors un motif de ce réseau et nous cherchons ses occurrences parmi plusieurs autres pièces de Ravel. Le résultat est montré fi-



FIG. 7.12: Motifs proposés par l'algorithme. Le motif de référence du troisième niveau, surligné, est représenté à droite de la figure.

figure 7.13. Les différentes occurrences de motif trouvées sont représentées successivement. Les valeurs de similarité pour chaque occurrence sont détaillées, ainsi que le nom du morceau duquel sont tirés les extraits. Une occurrence du motif recherché a été trouvée dans *Ma mère l'oie*, ce qui laisse imaginer les rapprochements possibles entre différentes pièces d'un même auteur ou d'auteurs différents que l'on aurait pas songé à rapprocher.

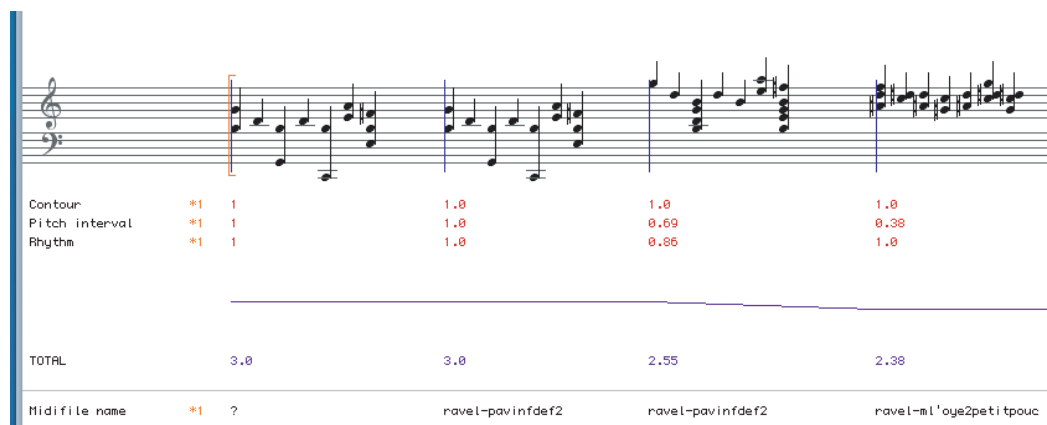


FIG. 7.13: Le résultat de la recherche du motif sélectionné figure 7.12. D'autres occurrences du motif ont été trouvées dans la suite de la Pavane, mais aussi dans *Ma mère l'oie*.

Chapitre 8

Conclusion

L'analyse musicale automatique est un domaine qui suscitait encore récemment plusieurs questions quant à sa légitimité et quant aux techniques à mettre en oeuvre pour l'explorer. Les outils que nous avons développés dans cette thèse apportent un point de vue original permettant la découverte de nouvelles structures échappant à l'analyse traditionnelle mais souvent pertinentes pour notre perception. Ces outils ont été développés en fonction des problèmes spécifiquement musicaux que soulevait la systématique de notre approche, à l'inverse d'autres études qui se focalisent davantage sur les traitements que sur la nature des données. En cherchant un modèle proche des données à traiter, nous avons pu à la fois apporter de nouvelles problématiques à l'analyse musicale et répondre à ces problématiques par la définition de nouveaux traitements spécifiquement adaptés.

8.1 Deux bibliothèques informatiques

Nous avons ainsi développé un extracteur automatique de rythme, reposant sur l'extraction de la pulsation et de la métrique, permettant un premier formatage des données. Cet extracteur se présente sous la forme d'une bibliothèque informatique accessible dans l'environnement Open-Music. La bibliothèque d'extraction rythmique, appelée 'OMKant', est accessible par l'intermédiaire

d'un éditeur musical présentant différents menus manipulables par l'utilisateur. Cette librairie a fait l'objet d'une documentation qui se trouve en annexe de ce document.

Dans un deuxième temps, nous appliquons aux données quantifiées un algorithme d'extraction des répétitions, puis des motifs, aboutissant à une segmentation automatique de la partition. Le résultat constitue une analyse de l'œuvre qui se veut perceptivement pertinente au regard de toutes les analyses possibles. Un dernier algorithme permet de rechercher les variations d'un motif donné à l'intérieur d'un corpus musical. Les motifs propres à une pièce donnée peuvent alors se révéler faire partie d'un plus large corpus de morceaux du même compositeur ou de compositeurs différents. Ces fonctionnalités ont elles aussi fait l'objet d'un développement sous forme de librairie Open-Music, appelée 'MusicMap', dont la documentation se trouve aussi en annexe.

8.2 Amélioration de l'existant

Plusieurs améliorations des outils que nous avons présenté sont directement envisageables sans qu'un investissement trop important ne soit requis.

En ce qui concerne l'extraction de la pulsation, la liste des pulsations possibles (l.p.p) pourrait être interprétée en termes de métrique de façon à affirmer ou infirmer les hypothèses de l'algorithme dédié à cette tâche.

L'algorithme d'extraction de la métrique fournit davantage d'informations qu'il n'en est utilisé par la suite. Les différentes métriques détectées par l'algorithme lors de son parcours chronologique de la séquence pourraient par exemple être interprétées en termes de changements de mesure.

Les similarités par contour et par intervalles présentent l'inconvénient d'être basées sur les seuls événements situés sur les pulsations, mais il devrait être possible de se servir de la similarité rythmique, basée sur tous les événements, pour sélectionner des événements intermédiaires aux pulsations qui soient présents dans les deux séquences comparées, ce que nous avons fait

lors de la phase appelée ‘appariement des onsets’.

Concernant le choix des corpus à analyser, il serait intéressant de se procurer plusieurs enregistrements MIDI de musiques atonales de façon à tester la généralité de nos choix.

Enfin, il devrait être possible de déduire du réseau de motifs trouvé comme étant le plus pertinent une première typologie liée à la segmentation de la séquence, en termes de phrases, développements, nouveaux éléments et reprises. Cette typologie pourrait être confrontée au jugement de plusieurs auditeurs, dans le cadre par exemple d’une campagne de tests en collaboration avec des psychoacousticiens, de manière à obtenir une évaluation plus systématique des résultats.

8.3 Travaux futurs

Nous abordons maintenant les travaux qui nécessiteraient un plus gros investissement, mais qui ne nous paraissent pas complètement irréalisables.

Tout d’abord, il faudrait étudier plus précisément certains styles rythmiques restant difficiles à analyser à cause de l’utilisation continue du swing ou des rubatos. De nouveaux marquages pourraient alors permettre une meilleure stabilité de l’algorithme. Par ailleurs, les grandes variations de pulsation ne sont actuellement pas interprétées par l’algorithme à l’échelle du morceau puisqu’elles sont utilisées essentiellement pour le régulariser, mais elles pourraient fournir des informations intéressantes pour l’analyse motivique. Nous pensons par exemple aux transitions entre mouvements lents et rapides. De même, certaines notions que nous avons utilisé pour l’extraction rythmique comme l’attente ou la modélisation des accentuations pourraient être formalisées dans notre système d’extraction de motifs, et inversement les informations de répétitions motiviques pourraient peut-être nous permettre de résoudre certains problèmes d’extraction rythmique.

L’extraction de motifs aboutit à une segmentation du morceau, mais de nombreuses informations contenues dans la matrice de similarité ne sont pas

utilisées. Toutes les répétitions du morceau ne sont pas détectées, et elles ne pourraient sans doute pas l'être par le même algorithme. Il reste à étudier lesquelles pourraient être trouvées moyennant une modification la plus légère possible des procédures d'analyse. Comment aussi trouver les répétitions les plus originales, c'est à dire celles qui sont perceptibles mais pas toujours de manière consciente. Les différents paramètres de contrôle pourraient peut-être être mis en relation avec un ou plusieurs effets perceptifs ce qui permettrait de les approcher de façon plus intuitive. Le fait de limiter notre étude de similarité aux événements situés sur les battues de pulsation pourrait, en plus de la piste proposée dans la section précédente, tirer parti de la redéfinition que nous faisons de la similarité pour l'étude des variations d'un motif donné dans un corpus. En effet, la diminution de la complexité combinatoire nous a permis d'expérimenter et d'ajuster notre modèle de matrices de similarité par rapport à la réalité musicale, mais il serait peut-être temps maintenant de revenir au problème de la complexité, par exemple en optimisant le code comme proposé dans [44] et en se servant de plus grosses machines. Le processeur utilisé dans tous nos exemples, un PowerPc G4 466Mhz, apparaît au bout de trois ans de thèse un peu lent !

La prise en compte de la polyphonie constitue à nos yeux un progrès important par rapport aux autres algorithmes proposés. Plus qu'un simple problème multidimensionnel, elle représente pour nous un enjeu spécifiquement musical, tout comme le temps constitue un élément fondamental et très particulier de la musique. De manière générale, nous pensons qu'il est nécessaire d'aller encore plus loin dans l'étude des différentes spécificités propres à la musique, car la complexité musicale à laquelle nous faisons face est loin d'être totalement maîtrisée. Si cet objectif pouvait être apparu avec plus d'évidence au cours de la lecture de cette thèse, les espoirs qu'y a placés l'auteur seraient déjà pour une part récompensés.

Annexes

Annexe A

Documentation du logiciel MusicMap

A.1 Introduction

Musicmap is a library for the analysis of repetitions in MIDI files. It relies on an algorithm that evaluates the similarity between two MIDI sequences by providing a number between 0 and 1. The results of an analysis are mainly represented in a similarity matrix with colored cells in order to visualize the different repetitions contained in the MIDI sequence. Two main functionalities are proposed :

- Automatic search and representation in a similarity matrix of some motivic repetitions and variations contained in a MIDI file
- Search for polyphonic patterns in any MIDI database that are similar to one that has been selected

The functionalities will be detailed below. A list of related papers at the end of the document provides more information on the algorithms that we employ. A tutorial is also available in OpenMusic.

✓ (a) Automatic search of repetitions
(e) Extract patterns from the matrix
(c) Compare the sequence in the orange brackets with...
(s) Save current analysis
(l) Load an analysis (.ana)
(0) View general matrix
(1) View contour matrix
(2) View pitch matrix
(3) View rhythm matrix
(p) Parameters
(shift) Stop the computation
(h) Help

FIG. A.1:

A.2 General description of the library

You can access to musicmap through its graphical editor in Open Music, called 'musicmap'. It is like a chord-seq (you can edit notes and play them) in which several functions have been added.

All of the functions that are specific to 'musicmap' can be activated from the menu called "musicmap" at the bottom right of the editor (see figure A.1).

The menu contains three main functions that are devoted to the analysis :

- a Automatic search of repetitions
- p Extract patterns from the matrix
- c Compare the sequence in the orange brackets with...

The other functions of the menu are devoted to edition, visualisation or saving of the results, loading of an old analysis or parameterization. They will be detailed along the document. The two following use-cases illustrate the two functionalities that we have mentionned in the introduction :

1. Load a MIDI file in the editor musicmap. Select the area in which you

want to make the analysis (see below How to select the area of the analysis). By default the area is the whole MIDI file. Ask for an automatic search of repetitions contained in the area that you have selected [(a) Automatic search of repetitions]. After the similarity matrix is computed, you can visualize salient structures in the matrix. By double-clicking on the cells you can visualize and listen to the patterns that correspond to them. You can also ask for an automatic extraction of the patterns that are judged as the most relevant by the algorithm [(e) Extract patterns from the matrix].

2. Load a MIDI file in the editor musicmap. Select the pattern that you want to compare (see below How to select the area of the analysis). Ask for a comparison between this pattern and potentially similar patterns contained in the same or other MIDI files [(c) Compare the sequence in the orange brackets with...]. After the computation, a new window appears that sequentially represents the patterns that have been found similar, sorted by decreasing similarity.

How to select the area of the analysis : it is possible to choose the area of the sequence that you want to analyse (use-case 1) or compare (use-case 2) by moving the orange brackets that are at the beginning and the end of the MIDI score (see figure A.2). You can move the orange brackets by using shortcuts ('alt'+ 'shift'+click for the delimiter at the beginning, or 'alt'+ 'shift'+click for the delmimitor at the end) or by dragging them with your mouse.

We will now detail the two main functionalities of the library.

A.3 Pattern extraction

A.3.1 Global description

Step 1 - Load a midifile

Once you have loaded a MIDI file in 'Musicmap', you can analyse it directly, but in this case an algorithm will first automatically extract the beat

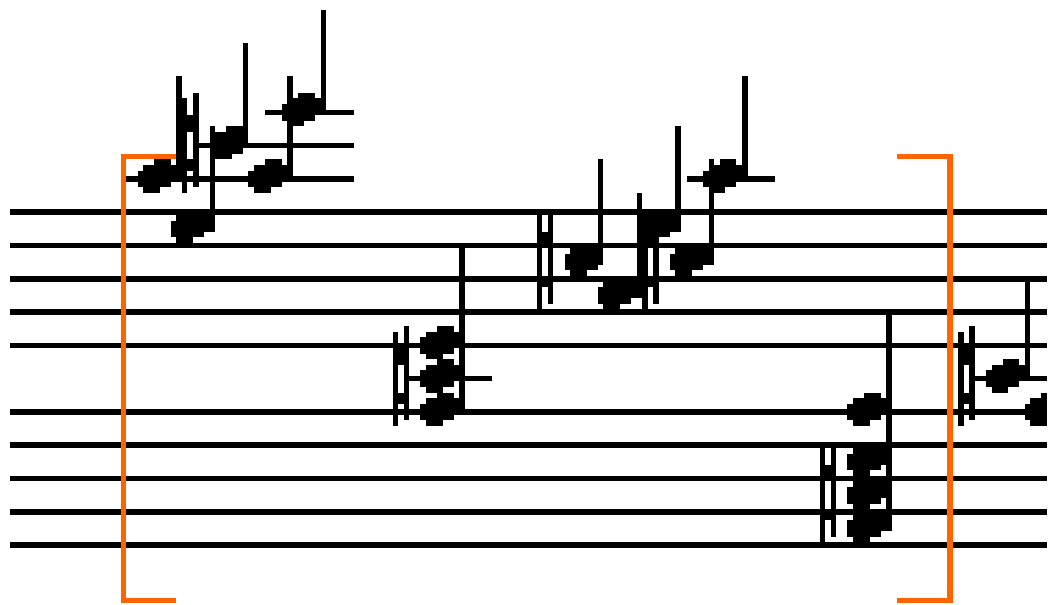


FIG. A.2:

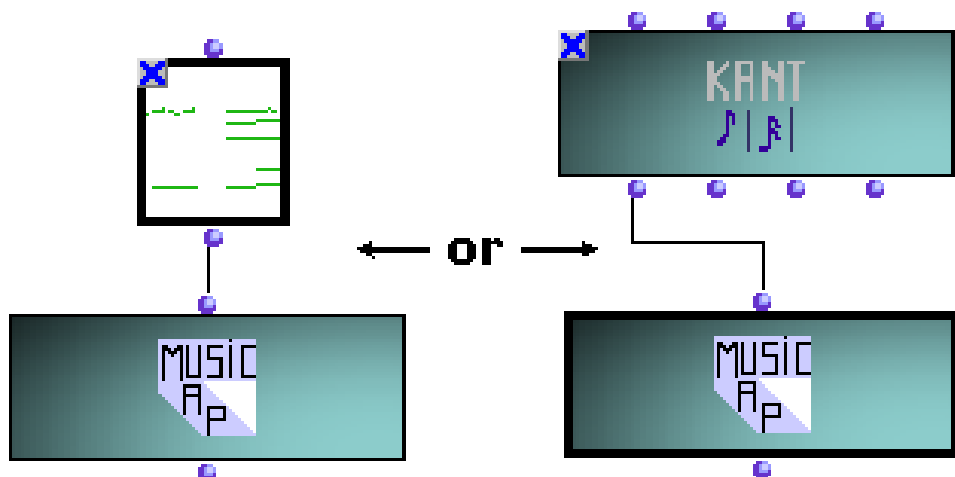


FIG. A.3:

that will be marked on the score by vertical segments. Indeed, the search for repetitions relies on the beats of the MIDI file that have to be defined before the analysis. Only the events that correspond to the beat positions will be analysed (see How and why do we track the beat of the sequence section). If you prefer that the algorithm relies on your own segmentation, you can use the librairie 'Kant' and then load the output of 'Kant' into 'Musicmap' (see figure A.3). Then, if you ask for automatic search of repetitions, the algorithm will rely on the segmentations that you have loaded. If you have loaded your segmentation from Kant but now you want automatic segmentation, delete the segmentations : double-click on one of the vertical blue lines and press '<-' (delete).

How to delete vertical blue lines : Double-click on one vertical blue line. Vertical blue lines between the orange brackets appear in black. Then press '<-' (delete). The vertical lines in black disappear. If now you choose 'Automatic search of repetitions', the algorithm will perform a new automatic segmentation of the sequence. If you do not delete the vertical blue lines, the algorithm will not make a new automatic segmentation but will use the

vertical lines on the score.

How and why do we track the beat of the sequence : the beat tracking algorithm is part of the openmusic library OMKant 3.1. More details on the algorithm can be found in the OMKant documentation or in the related papers that we give at the end of the document. If you want to specify your own segmentation of the sequence, you can use OMKant 3.1 and then connect its output to musicmap.

Tracking the beat is a necessary step because the time computation required for comparing all the events (not only the beats) of all possible pairs of all possible sequences contained in a MIDI file would be far too long. The first way to reduce the time computation consists in comparing only sequences of same length (it also solves the issue of how to match the events of sequences that have different length : one event of one sequence would not automatically correspond to an event at the same position in the other sequence). Then we have to define what is the length of a sequence. We could define it as the number of events of the sequence, but still too many sequences could be matched and the time computation would be high.

Our solution is to consider sequences that have same number of beats. Doing that, if variations can occur between the beats the sequences would still be judged as the same. Moreover, the number of sequences that will be compared is highly reduced. On the figure A.4, the beat positions are defined by the blue vertical lines. If you do not load your own segmentation from Kant, the algorithm will automatically segment the sequence before analysing it.

Step 2 - Automatic search of repetitions

Select in themusicmap menu 'Automatic search of repetitions' (or press a). The search for repetition consists in comparing all the sequences of the MIDI file that have the same number of beats. The result of the comparison between each pair of sequences is represented in a similarity matrix. You can

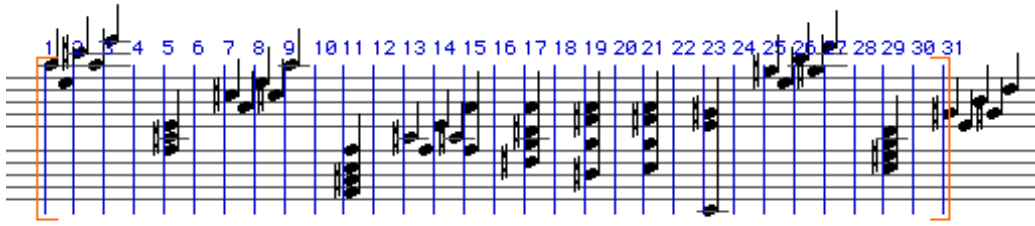


FIG. A.4:

stop the computation by pressing 'shift' if you are satisfied with the partial matrix that progressively appears under the score.

How to choose the length of the sequences that are compared : you can define the number of beats of the sequences that will be compared by selecting the box "number of beats of the sequences that are compared" in the musicmap menu "parameters" (see figure A.5).

Step 3 - Extract patterns from the matrix

Finally, you can select the function 'extract Patterns from the matrix' (or press e). This algorithm automatically extracts patterns that have the longest size or that are the most often repeated. After the computation, a new window appears that contains two open-music 'maquettes'. The maquette at the bottom contains the three longest patterns, the maquette at the top contains the possible subdivisions of the longest pattern. Each maquette must be read from bottom (long patterns) to top (short patterns). More details are given below in 'How do we extract patterns from the matrix'.

How to save and load an analysis : It is possible to save an analysis, and then to load it from default directory "Music-Map0.1/Analyses". You can use either shortcuts ('s' for saving and 'l' for loading) or the commands "save current analysis" and "load an analysis (.ana)" in the musicmap menu.

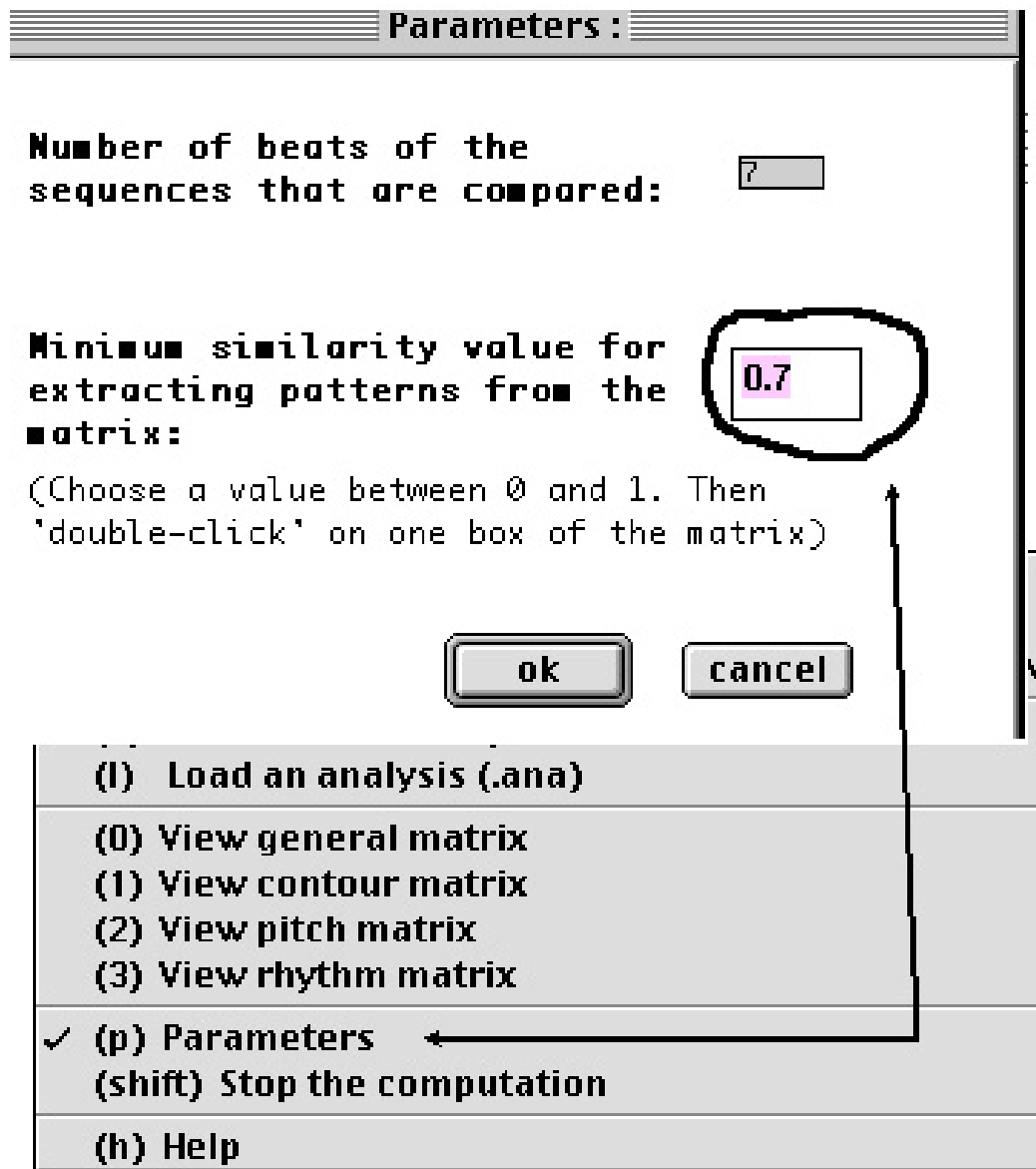


FIG. A.5:

A.3.2 Automatic search of repetition

How do we compute the similarity between two sequences

The similarity measure provides a value between 0 and 1 for each comparison between two musical sequences that have same number of beats. 0 is for sequences that do not appear as similar (to the algorithm), and 1 is for sequences that appear as identical. We do not measure dissimilarity.

What are 'corresponding beat segments' : we only compare sequences that have same number of beats, so each beat position of one sequence correspond to one beat position of the other sequence. All our similarity measures rely on this assumption. We will call corresponding beat segments the events between two consecutive beats of one sequence that are associated to the events between the same beat positions in the other sequence.

The similarity value is the average of three values between 0 and 1 corresponding to three viewpoints :

- 1 the contour (horizontal intervals at the top and bottom of the sequence)
- 2 the pitch intervals (vertical and horizontal sets of intervals inside the polyphony)
- 3 the rhythm

Parties non détaillées :

- *Cognitive aspects of the similarity measure*
- *the contour*
- *the pitch intervals*
- *the rhythm*

Graphical representations of the results

What can we read in the similarity matrix :

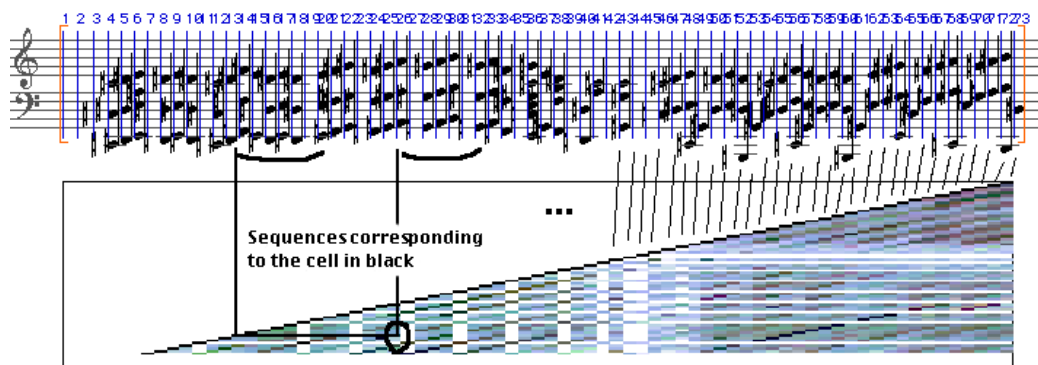


FIG. A.6:

Each cell of the matrix corresponds to the comparison between two MIDI sequences of same length (see How to choose the length of the sequences that are compared in step 2 for changing the default length). From left to right and from bottom to up, the coordinates of each cell indicate the beginning positions of the two corresponding compared sequences. Both vertical and horizontal axis can be seen as temporal grids whose units are the beats of the MIDI file. For instance, the black diagonal at the top of the matrix corresponds to sequences that are compared to themselves. The line at the bottom of the matrix corresponds to the comparison of the beginning of the MIDI file with successively all the other sequences of same length contained in the MIDI file (see figure A.6).

If you double-click on one cell, the corresponding sequences appear in a new window. Each cell has a color that corresponds to the similarity value between the two corresponding sequences. The darker the color of the cell, the higher the similarity. Once you have doubleclicked, two windows appear.

The first one contains the compared sequences. Red lines correspond to the real compared sequences, blue lines are the continuation of the sequences proposed by the algorithm (you can choose the threshold value, see below How to choose the similarity threshold for visualizing patterns in the matrix). The second window contains only the part of the sequences that has been

used to compute the similarity value between the two sequences. It can be seen as the template of the sequences. We will describe this notion together with the description of the similarity measure (see what are the templates). Diagonals (from bottom left to top right) of dark color in the matrix indicate that long sequences, whose size is higher than the one that was chosen for the computation of the cells, are similar. How to choose the similarity threshold for visualizing patterns in the matrix. It is possible to visualise patterns corresponding to long sequences by choosing the minimum similarity value of the consecutive cells in the diagonal that are part of the pattern. For that, choose a value in the box "Minimum similarity value for extracting patterns from the matrix" after selecting "Parameters" in the menu "Musicmap" (see figure A.5).

What are the templates :

The templates appear in the second window when you click on one cell of the matrix. The two sequences of the window correspond to the events of the compared sequences that have been used to establish the similarity. A template is composed of three types of events corresponding to the three similarity measures. In order to distinguish the events according to the different similarity measures, we have given different channels numbers to them. Channel number 2 (red) is for contours, channel number 1 (blue) is for intervals and channel number 10 (grey) is for rhythm. For rhythm, only onsets are played (pitches of the corresponding events do not play any roles). You can listen to the template to evaluate the difference between the initial sequence and what has been considered by the computer for the similarity.

Where do the similarity values appear in the editor :

When double-clicking on one cell of the matrix, the similarity value between the corresponding sequences appears in the listener window. It is detailed as below :

The last number (for instance number 6 in figure A.7) corresponds to the exact length of the sequences that has been used for the comparison, starting from the last event of the sequence. You can also visualize separately the



FIG. A.7:

different similarity values (see 19).

How to visualize the different matrices corresponding to the different similarity values :

The matrix you visualize is computed from three other matrices that correspond to the different viewpoints we adopt to compute the global similarity (contour, pitch intervals and rhythm). You can visualize each matrix corresponding to each viewpoint by selecting in the menu "musicmap" either "view contour matrix" (or press 1), "view pitch matrix" (or press 2) or "view rhythm matrix" (or press 3).

A.3.3 Extracting patterns from the matrix

We propose an algorithm that automatically extracts two kinds of patterns after having analysed the sequence (press 'a' for analyse and 'e' for extracting patterns) : long patterns and the different components of the longest pattern. Note : that patterns that are extracted can cover themselves. The long patterns are represented in an Openmusic 'maquette'. Three patterns are represented in the maquette. They are sorted by size from the bottom to the top of the maquette. Each pattern is associated with its repetition in the MIDi file, and is represented in a chronological order on the horizontal axis of the maquette.

The components of a long pattern are also represented in an Openmusic 'maquette'. They could be seen as the cells or motives of a musical phrase (assuming that the long pattern is a phrase). They are represented in the maquette from bottom to top, from the long pattern to the shortest pattern

contained in it.

A.4 Search for a polyphonic pattern in a MIDI database

A.4.1 Global description

You can proceed as follows :

1. Load a midifile and select the motive you want to compare :

Once you have loaded the MIDI file, you can use the orange brackets to select the beginning and end of the motif that you want to compare (see How to select the area of the analysis in part I for the procedure).

2. Press 'c' or select in the musicmap menu "(c) Compare the sequence in the orange brackets with..." :

A new window appears in which you can choose the MIDI files in which you want to make the search (see figure A.8). If you select "current sequence", the algorithm will make the search in the score in which you selected the motive to compare. If you choose "one MIDI file", the algorithm will ask you to choose a MIDI file in which it will make the search. If you choose "several MIDI files", the algorithm will ask you a repertory instead of a file.

Then, during the computation, the numbers that appear in the listener correspond to the current position in the sequence of the analysis (in milliseconds). If the computation takes too much time, you can press the key 'shift' in order to stop the computation and see the current results. Then, the result of the computation appears in a new window, in which the sequences similar to the one that you chose to compare are sorted according to their decreasing similarity value. The different similarity values are detailed in columns (respectively contour, pitch interval and rhythm) and the total value appears under the graph. If you made the search in a database, the name of the corresponding midifile for each pattern appears at the bottom in the line called "Midifile name". You can try to play the sequence and compare the

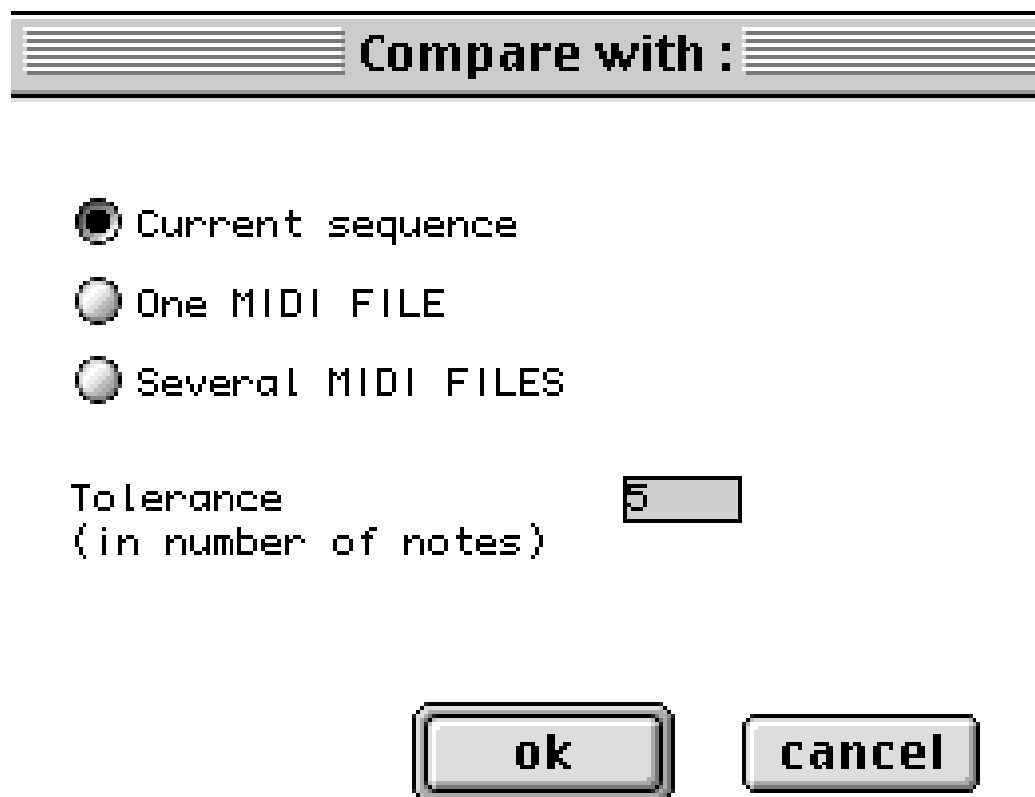


FIG. A.8:

results with your own judgement on the similarities.

A.4.2 More details on the similarity measure

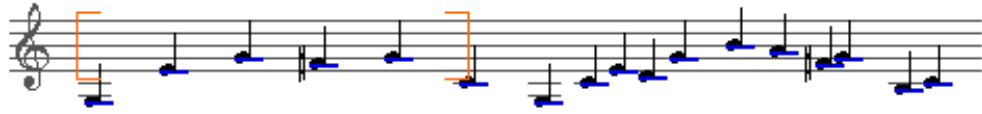
The similarity measure we employ is very similar to the one that we use for automatic extraction and representation of patterns. However, we do not rely anymore on a segmentation of the MIDI score. Indeed, as few sequences are compared, it is possible to evaluate the similarity between the motive that has been selected and all the possible sequences that have same number of eventually non consecutive events contained in the list of MIDI files in which you want to make the search.

Thus, sequences of different temporal length can be matched. You can choose a tolerance value that determines the maximum difference between the number of onsets of each sequence (see figure A.8). This number is initialised at five. It means that if a sequence contains six onsets, it will not be compared to sequences that contain more than 11 onsets ($6 + 5 = 11$). A constraint has been implemented in order to decrease the similarity value for sequences that do not have identical proportional time relationship.

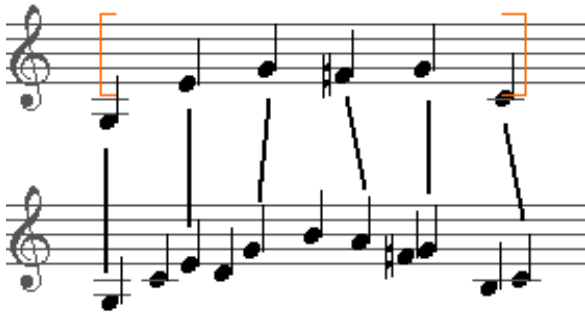
Example :

For instance, let's consider the sequence in figure A.9. In the window 1, the pattern that we want to search has been delimited by the orange brackets. We then ask the algorithm to search for this pattern in the whole sequence. Of course, the algorithm retrieve the original pattern and sort it as the most similar to itself, but in second position, the algorithm extracts the sequence that is represented at the bottom of window 2 and 3. This sequence is not identical but contains similar pitches and similar onsets. Before extracting this sequence, the algorithm has computed all the possible matches between the original pattern and all the possible patterns of same number of notes contained in the sequence in window 1. This result illustrates how sequences with different pitches and radically different beat positions but similar onsets can be matched. In window 2, we represent the matching pitch by pitch between sequence at top and sequence at bottom. We can observe quite

1



2



3

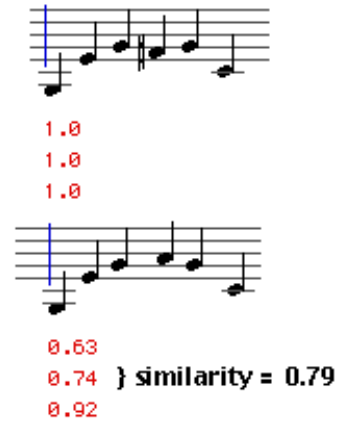


FIG. A.9:

different onsets positions but similar proportional ratios between the two matched sequences. In window 3, we represent the different similarity values that have been computed for the matching. Contour is valued 0.63, melodic line is valued 0.74 and rhythm is valued 0.92. Total similarity is 0.79. Pitches that were positionned between the matched events are not represented.

Bibliographie

- [1] Agon Carlos, (1998) *Openmusic, un langage visuel pour la composition musicale assistée par ordinateur* Thesis, Ircam
- [2] Assayag Gérard, Rodet Xavier (2002) *Rapport sur les travaux du groupe Outils pour l'analyse musicale* Ircam
- [3] Bent Ian, (1997) *L'analyse musicale - histoires et méthodes* éditions main d'œuvre, collection musique et mémoire
- [4] Boulez Pierre, (1964) *Penser la musique aujourd'hui* Paris, Gonthier
- [5] Brown Judith, (1993) *Determination of the meter of musical scores by autocorrelation* Journal of the Acoustical Society of America 94 (4) 1953-1957
- [6] Cambouropoulos Emilios, (1998) *Towards a General Computational Theory of Musical Structure* PhD, Edinburgh, Faculty of Music and Department of Artificial Intelligence
- [7] Cemgil, Kappen, Desain, Honing (2000) *On tempo tracking : Tempogram representation and Kalman filtering* proc.ICMC p.352-355
- [8] Clausen, Engelbrecht, Meyer, Schmitz (2000) *PROMS : a web-based tool for searching in polyphonic music* proc.Ismir 2000
- [9] Conklin D, Witten I.H (1995) *Multiple viewpoint systems for music prediction* Journal of new music research, 24 :51-73
- [10] Cope David, (1993) *A computer model of music composition* Machine models of music, S.M Schwanauer et D.A Levitt (eds), MIT press, Cambridge (Ma)

- [11] Cooper et Meyer, (1960) *The rhythmic structure of music* University of Chicago Press, Chicago, Illinois
- [12] Desain et Siebe de Vos, (1990) *Auto-correlation and the study of Musical Expression* Proceedings of the International Computer Music Conference. 357-360.
- [13] Desain et Honing, (1989) *Quantization of musical time : a connectionist approach* Computer Music Journal 13(3)
- [14] Forum Diderot, (1999) *Logiques mathématiques, logiques musicales au XXème siècle* Quatrième forum mathématique Diderot, Paris, Ircam
- [15] Dixon Simon, Emiliós Cambouropoulos (2000) *Beat tracking with musical knowledge* proc. ECAI
- [16] Dixon Simon, (2001) *Automatic Extraction of Tempo and Beat from Expressive Performances* Journal of New Music Research, 30, 1, pp 39-58
- [17] Dixon Simon, (2001) *An empirical Comparison of tempo trackers* 8th Brazilian Symposium on Computer Music, Fortaleza, Brazil
- [18] Forte A, (1989) *La Set-complex theory : élevons les enjeux!* Analyse musicale, 4e trimestre, pp. 80-86
- [19] Lachenmann Helmut (1996) *Zum Problem des Strukturalismus, Musik als existentielle Erfahrung* par Joseph Häusler, BreitkopfHärtel
- [20] Large E et Kolen J (1994) *Resonance and the Perception of musical meter* Connection Science, 6 (1) 177-208
- [21] Lartillot Olivier, (2003) *Perception-Based Advanced Description of Abstract Musical Content* Proc. WIAMIS, London
- [22] Lerdahl et Jackendoff, (1983) *A generative theory of tonal music* Cambridge, MA : MIT Press
- [23] Levy Fabien, (1995) *La complexité du processus d'écriture musicale* Mémoire de dea, Ircam - EHESS - CNRS

- [24] Lidov David, (1975) *On musical phrase* Groupe de recherches en sémiologie musicale, faculté de musique, université de Montréal, Montréal, Canada
- [25] Longuet-Higgins, (1987) *Mental process* Cambridge, Massachussets, MIT press
- [26] Lusson Pierre (1986) *Place d'une théorie générale du rythme parmi les théories analytiques contemporaines* Analyse Musicale n°2, pp. 44-51, 1986.
- [27] Malt, Mikhail (1998) *Modèles mathématiques et composition assistée par ordinateur* Thèse, EHESS - Ircam, Paris
- [28] Mazzola Guerino (2002) *The topos of music* Chap 39, Birkhauser Verlag, Basel-Boston-Berlin
- [29] Meredith David, Kjell Lemström, Geraint Wiggins, (2003) *Algorithms for discovering repeated patterns in multidimensional representations of polyphonic music* Cambridge Music Processing Colloquium, Department of Engineering, University of Cambridge.
- [30] Meudic Benoit, St-James Emmanuel (2003) *Automatic Extraction of Approximate Repetitions in Polyphonic Midi Files Based on Perceptive Criteria* in Lecture notes in Computer science, LNCS 2771 Springer Verlag
- [31] Meudic Benoit, (2003) *Musical similarity in polyphonic context : a model outside time* Proceedings of XIV CIM (Colloquium on musical informatics)- Firenze, Italy
- [32] Meudic Benoit, (2002) *A causal algorithm for beat tracking* 2nd conference on understanding and creating music, Caserta, Italy
- [33] Meudic Benoit, (2002) *Automatic meter extraction from midifiles* Proc. JIM, Marseille
- [34] Meyer Leonard, (1973) *Explaining music : essays and explorations* Berkeley, R 1978
- [35] Nattiez Jean-Jacques, (1975) *Fondements d'une sémiologie de la musique* Paris

- [36] Nattiez Jean-Jacques, (1987) *Musicologie générale et sémiologie* p.313, Paris
- [37] Narmour Eugène, (1982) *Towards an analytical symbology : the melodic, harmonic and durational functions of implication and realization* Musical grammar and computer analysis : modena 83-114
- [38] Peeters Geoffroy, Amaury La Burthe, Xavier Rodet , (2002) *Toward Automatic Music Audio Summary Generation from Signal Analysis* ISMIR, Paris, France
- [39] Rameau J-P, (1722) *Traité de l'harmonie réduite à ses principes naturels* Paris
- [40] Réti R, (1951) *The thematic process in music* New-York
- [41] Riemann, H , (1903) *System der musikalischen rhythmik und metrik* Leipzig, Breitkopf und Hartel
- [42] Rispoli Raphaël, (2002) *Eléments d'une méthodologie computationnelle pour la construction autonome des représentations. Application à l'élaboration de descriptions avec le système FADA* Université Paris6
- [43] Rivette Jacques, (1959) à propos de la sortie d'Hiroshima mon Amour d'Alain Resnais Les Cahiers du Cinéma, numéro de juillet
- [44] Rolland Pierre-Yves, (1999) *Discovering patterns in musical sequences* JNMR 28 N°4 pp. 334-350
- [45] Rousseau Jean-Jacques, (1781) *Essai sur l'origine des langues* œuvre posthume, éd A.Belin, Paris
- [46] Ruwet N, (1966) *Méthodes d'analyse en musicologie* RBM, XX, 65, repr dans *Langage, musique et poésie* Paris, 1972, 100
- [47] Scheirer Eric, (2000) *Music-Listening Systems* Ph.D. dissertation, MIT Media Laboratory
- [48] Schenker Henrich, (1935) *Neue musikalische theorien und phantasien* Vienne, trad.fr de Nicolas Meeùs *L'écriture libre* Liège
- [49] Temperley David, (2001) *The cognition of basic musical structures* Massachusetts institute of technology

- [50] Tovey Donald Francis, (1949) *Essays and lectures on music* éd H.J Foss
- [51] Tversky A, (1977) *Features of similarity* journal of Psychological Review,
p327-352