

Chapitre 9

Algorithmes et techniques pour l'analyse musicale

9.1. Introduction

L'interprétation d'un phénomène musical complexe suppose d'être capable, consciemment ou non, de le décomposer et d'établir des liens entre ses éléments. La notion de *structure musicale* est prise ici dans le sens de l'organisation attribuée à *une surface musicale*¹ en termes de parties et de relations/fonctions entre elles à divers niveaux de description. L'analyse musicale vise la révélation de ces descriptions structurelles, souvent d'un point de vue perceptif/cognitif (Plié, 1980).

On emploie souvent des méthodologies informatiques pour étudier les processus analytiques musicaux *en eux-mêmes* et également pour construire des outils d'analyse utilisables dans des applications musicales pratiques comme l'exécution interactive, la composition assistée par ordinateur, la recherche et l'indexation de données musicales, l'expressivité de l'exécution automatique, etc. Dans ce dernier cas, le principal argument est que plus on vise des applications informatiques musicales sophistiquées, plus le système informatique doit « comprendre » les structures musicales.

Chapitre rédigé par Emílios CAMBOUROPOULOS et Pierre-Yves ROLLAND.

1. La surface musicale est considérée comme « le plus bas niveau de représentation ayant une signification musicale » [JAC 87, p. 219]. Par commodité, elle est considérée ici comme une séquence de notes discrètes représentées de façon symbolique, sans éléments structurels de plus haut niveau (marques d'articulation, signature rythmique, etc.)

Il existe différents points de vue sur la façon de créer des modèles informatiques pour les tâches musicales. Une distinction entre les stratégies de modélisation est basée sur la quantité de *connaissances* explicites directement intégrée dans la représentation informatique [CON 95] : dans l'approche relevant de l'*ingénierie des connaissances*, la représentation est entièrement codée « manuellement » par le programmeur-théoricien sur la base de connaissances musicales intuitives ou explicites (par exemple dans les manuels d'enseignement de la musique), alors que dans l'approche relevant de l'*induction empirique*, la représentation est élaborée à partir de généralisations sur un ensemble de phénomènes musicaux d'après divers principes fondamentaux généraux et diverses stratégies. Du fait de la diversité de styles et de « langages » de la musique, les chercheurs ont souvent eu recours à la première approche, avec laquelle ils ont construit des systèmes experts complexes pour effectuer des tâches analytiques particulières dans des domaines musicaux spécialisés (par exemple le contrepoint du XVI^e siècle dans le style de Palestrina, l'harmonie tonale du XVIII^e siècle, l'analyse schenkérienne, l'analyse timbrale du XX^e siècle pour la musique atonale, l'harmonie du jazz, des systèmes correspondant à divers genres musicaux traditionnels, etc.). D'autre part, certains chercheurs ont essayé de construire des systèmes inductifs et moins dépendants de styles musicaux spécifiques ; ces approches utilisent souvent des techniques d'apprentissage automatiques, des méthodologies relevant de la théorie de l'information, des principes et modèles cognitifs, etc. Ce chapitre donne une description brève des modèles relevant de la deuxième stratégie, car ils sont habituellement plus généraux et peuvent être appliqués à des styles et idiomes musicaux variés.

Divers modèles informatiques pour des tâches analytiques *mélodiques* seront décrits par une série d'exemples musicaux. La plupart de ces techniques peuvent être adaptées et appliquées aux fichiers musicaux polyphoniques en tenant compte de paramètres complémentaires (par exemple l'harmonie : on considère un passage musical comme une série d'accords) ou si l'on dispose d'algorithmes de séparation de voix permettant de décomposer un passage polyphonique en plusieurs « mélodies » indépendantes (voir l'aperçu donné dans [TEM 01]).

En supposant que la mélodique est présentée sous la forme d'une séquence de notes ayant des valeurs nominales, les modèles informatiques présentés ici répondent aux questions suivantes :

- a) comment détecter les frontières locales pour commencer à segmenter une mélodie ?
- b) comment détecter les structures accentuelles et métriques ?
- c) comment détecter des patterns musicaux significatifs ?
- d) quel est l'effet de la similitude musicale sur la segmentation mélodique ?
- e) comment organiser les segments musicaux en catégories « significatives » (par exemple motifs et thèmes) ?

Ce chapitre a pour but de présenter des stratégies permettant de résoudre ces problèmes d'analyse musicale, non de signaler les meilleurs algorithmes (il existe peu d'algorithmes suffisamment testés et peu d'études comparatives, car la musicologie informatique en est toujours à un stade peu avancé). Pour illustrer le propos de façon simple, il illustre l'application d'une série d'algorithmes d'analyse à une mélodie simple, mentionne d'autres méthodes possibles et fournit quelques exemples supplémentaires.

9.2. Frontières locales

Les principes de la Théorie de la forme (*Gestalttheorie*) relativement à l'organisation de la perception, sont un ensemble de règles pratiques portant sur les modes préférentiels de groupement des événements (notamment visuels) en schémas de plus grande échelle. Selon deux des principes de *Gestalt* relativement aux données de bas niveau, les objets proches (principe de Proximité) ou similaires (principe de similitude) tendent à être perçus comme des groupes. Ces principes sont à la base de divers modèles récents de détection des frontières locales dans les séquences mélodiques.

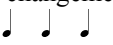
Par exemple, dans le modèle de Tenney et Polansky [TEN 80], les principes de Proximité et de Similitude sont pris comme des descriptions différentes du même phénomène, à savoir un maximum local de la distance entre événements musicaux consécutifs pour n'importe quel paramètre musical, par exemple la hauteur et l'instant initial de la note. Une mélodie donnée est représentée par une séquence d'intervalles de hauteur (en demi-tons) et par une séquence d'intervalles entre instants initiaux (exprimés en multiples de la plus petite unité de durée adéquate ; dans l'exemple ci-dessous il s'agit de doubles-croches). On calcule alors la somme de chaque paire correspondante de données dans les deux séquences de valeurs, et les maximums locaux de la série de sommes sont considérés comme les frontières les plus probables (figure 9.1b).

Le modèle de règles de regroupement locales de Lerdahl et Jackendoff [LER 83] essaie lui aussi de repérer les maximums locaux dans les séquences d'intervalles paramétriques. Ce modèle propose des règles plus variées, mais elles sont binaires (aucune valeur ne leur est associée), par exemple :

- *GRP2b* : un intervalle entre instants initiaux est perçu comme une frontière locale s'il est plus grand que le précédent et que le suivant ;
- *GPR3a* : un intervalle de hauteur est perçu comme une frontière locale s'il est plus grand que le précédent et que le suivant ;

– *GPR3d* : dans une séquence de quatre notes, on perçoit une frontière entre les deux notes centrales **si les deux premières et les deux dernières notes ont la même durée si les deux notes du milieu ont des durées différentes.**

Une règle complémentaire *d'intensification* indique que si plusieurs règles de détail locales s'appliquent dans une position donnée, la frontière locale correspondante est plus marquée. Un exemple est donné dans la figure 9.1a (où seules interviennent les trois règles ci-dessus). Ce modèle a été très influent ; beaucoup d'applications informatiques ont adopté et adapté les règles de Lerdahl et Jackendoff pour la segmentation (par exemple [TEM 01]). Ce modèle, évalué expérimentalement, s'est avéré très fiable [DEL 87].

Selon Cambouropoulos [CAM 96, CAM 97], bien que cette formalisation des principes de la Gestalttheorie (un grand intervalle situé entre de petits intervalles) fournisse le principal facteur de découverte des frontières locales, une approche plus générale se devrait de représenter *n'importe quel* changement d'échelle des intervalles. Par exemple, dans la séquence de durées , un auditeur entend facilement un point possible de segmentation alors qu'aucun des deux modèles précédents ne suggère de frontière.

Le modèle de segmentation que nous proposons et que nous appelons modèle de détection de frontières locales (MDFL, en anglais *Local Boundary Detection Model* *LBDM*) est basé sur deux règles : la règle de changement d'identité et la règle de proximité. La règle de changement d'identité est plus simple que tous les principes de la Gestalttheorie, car elle est applicable à partir de deux entités (c'est-à-dire qu'elle permet de juger si deux entités sont identiques ou non), tandis que la règle de proximité exige au moins trois entités (deux entités sont jugées plus proches ou plus similaires que deux autres entités) :

– *règle de changement d'identité (RCI)* : les frontières peuvent être présentées sur n'importe lequel de deux intervalles consécutifs si ces intervalles sont différents. Si les deux intervalles sont identiques, aucune frontière n'est suggérée ;

– *règle de proximité (RP)* : si deux intervalles consécutifs sont différents, la frontière présentée sur le plus grand intervalle est proportionnellement plus forte.

Le modèle de détection de frontières locales suggère toutes les positions possibles des frontières locales sur une surface musicale, avec divers degrés de relief. Les valeurs sont normalisées sur l'intervalle [0,1]. Les sommets locaux sont les points les plus susceptibles d'être perçus comme des frontières locales (voir l'exemple de la figure 9.1a et la description formelle de l'algorithme en annexe, A.1). Une description détaillée de l'algorithme est donnée dans [CAM 01a]. Le *MDFL*, appliqué à de nombreuses mélodies, a fourni de très bons résultats. Une évaluation de ce modèle est présentée dans [BAT 98 ; HOT 02].

Nous nous bornerons à exposer une nouvelle hypothèse : que le regroupement local et les structures d'accentuation locales n'ont pas à être des composants indépendants d'une théorie du rythme musical, mais qu'ils sont étroitement associés, de telle façon que si l'un est défini, l'autre peut en être déduit automatiquement [CAM 97]. Cette hypothèse se base sur l'observation que les frontières de regroupement sont étroitement liées aux événements accentués entre lesquelles elles se situent. Une frontière perçue dans un continuum donné indique que les éléments qui le délimitent sont plus saillants que d'autres événements situés plus loin. Epstein écrit : « En fait, la démarcation revient à l'accent : c'est l'accent exigé à ce moment pour tracer la frontière d'un segment temporel » [EPS 95, p. 24].

Par exemple, en attribuant des scores de frontière locale (par exemple, des valeurs *MDFL*) à toutes les paires d'intervalles successifs, on obtient la structure d'accentuation locale de la surface. Les valeurs de cette séquence de niveaux d'accentuation indiquent la prégnance relative des notes au niveau local.



Figure 9.3. Niveaux d'accentuation locaux pour la mélodie initiale de la Valse de Chopin op. 18

La forte corrélation étroite entre le regroupement et les structures d'accentuation est importante car elle permet d'élaborer un modèle qui n'a pas besoin de deux méthodes distinctes pour la détection des frontières locales et pour les phénomènes d'accentuation. Contrairement au modèle de Lerdahl et Jackendoff (figure 9.4a), notre modèle relie directement la structure d'accentuation locale à la structure de regroupement (figure 9.4b), ce qui permet de créer plus facilement et plus efficacement des systèmes informatiques de traitement du rythme musical.

- | |
|--|
| <p>a. Structure de regroupement/structure métrique/
structure d'accentuation.</p> <p>b. Structure de regroupement/structure
d'accentuation ↔ structure métrique.</p> |
|--|

Figure 9.4. a) La théorie du rythme de Lerdahl et Jackendoff ; b) notre théorie

Une fois que la structure d'accentuation locale a été définie, on peut essayer de lui faire correspondre une structure métrique bien formée ; on trouvera des détails sur les modèles d'appariement de patterns dans [PAR 94]. Cet appariement peut porter sur plusieurs niveaux métriques hiérarchisés de battements rythmiques, ou sur un niveau unique, voire, probablement, le « niveau zéro », selon le type de musique.

Le score total d'accentuation, correspondant à un réseau métrique donné, peut être calculé en ajoutant simplement les accents de tous les événements dont le début coïncide avec les points du réseau. Si entre les divers(es) positions/déplacements d'un réseau métrique on trouve une valeur totale « significativement » plus grande, on considère que c'est le meilleur choix. Si les divers placements d'un réseau reçoivent des valeurs similaires, on peut penser qu'il y a ambiguïté métrique dans le réseau considéré. Par exemple, on peut tenter de faire correspondre des réseaux métriques en 2/4 et en 3/4 avec les accents de l'exemple mélodique.

Dans cet exemple, on suppose que le score d'accentuation de la première note est de 1. La somme de tous les accents de chaque réseau est telle que les sommes sont comparables. Le tableau 9.1 donne la valeur totale pour chaque déplacement de chaque réseau, multipliée par la valeur de réseau métrique lui-même * (pour que les sommes soient comparables). La meilleure correspondance est obtenue pour le réseau métrique 3/4 qui commence sur la première note de la mélodie. La valeur totale est calculée comme suit :

$$(1,00 + 0,95 + 1,13 + 1,35 + 0,44 + 0,39 + 0,39 + 0,39) \times 3/4 = 4,53$$

Ce réseau métrique est justement le rythme indiqué par le compositeur dans la partition.

1** La structure métrique peut être déduite de la structure d'accentuation, mais, en même temps, elle influence fortement et résout l'ambiguïté de la structure de groupement/accentuation. Des accents métriques peuvent être ajoutés aux niveaux d'accentuation de façon à obtenir une régulation de la structure de groupement d'un morceau.

L'hypothèse de bas niveau présentée ici ne couvre que les traits structuraux de bas niveau de l'organisation en termes d'accentuation et de groupement. Il est possible qu'aux niveaux plus élevés d'organisation ces structures soient partiellement indépendantes, voire en conflit. Il est en tous cas très intéressant de

1. Le mètre n'est pas simplement un phénomène mental suscité par la musique. Il a un statut psychologique autonome dans un contexte culturel donné et il influence la façon dont la musique est exécutée et perçue. Clarke décrit une expérience qui illustre l'influence de différentes structures métriques sur l'exécution d'une même mélodie [CLA 85].

voir tout ce que représente et permet d'inférer une structure de groupement locale bien définie (par opposition aux structures d'accentuation et métriques locales considérées indépendamment).

Réseau métrique	2/4		3/4		
Décalage	0	1/4	0	1/4	2/4
Valeur totale	3,01	3,95	<u>4,53</u>	2,39	3,11

Tableau 9.1. Découverte de la structure métrique préférée (valeur soulignée) pour le profil d'accentuation de la

9.4. Patterns musicaux

Un des principaux facteurs de compréhensibilité de la musique est l'autoréférence, c'est-à-dire le réseau de relations de nouveaux passages musicaux avec ce que l'on a déjà entendu. La répétition structurelle et la similitude sont des dispositifs cruciaux pour l'établissement de ces relations. Les entités musicales similaires sont organisées en catégories musicales comme motifs rythmiques et mélodiques, thèmes et variations, groupes de progression harmonique, etc. (section 9.6). La similitude musicale établit des rapports entre entités musicales différentes, mais permet en premier lieu de définir ces entités en contribuant directement à la segmentation d'une surface musicale en unités significatives (section 9.5).

On emploie souvent les techniques de traitement de patterns pour déterminer la similitude musicale. On fait, en général, l'hypothèse qu'il est possible de considérer une surface musicale comme une chaîne d'entités musicales (notes, accords, etc.) auxquelles on peut appliquer des techniques de reconnaissance de formes ou d'induction. Dans ce chapitre, le terme d'*induction de patterns* désigne les techniques qui permettent l'extraction des patterns utiles d'une chaîne, et celui de *reconnaissance de formes* renvoie aux techniques servant à trouver tous les cas d'un pattern prédéterminé dans une chaîne donnée. Les deux algorithmes d'induction de patterns indiqués aux paragraphes 9.4.1 et 9.4.2 peuvent être utilisés pour l'analyse mélodique automatisée. On trouvera des aperçus de l'application des algorithmes de traitement de patterns aux chaînes musicales dans [CAM 99, CRA 98, ROL 99].

****note**

2. Il existe de nombreux algorithmes d'appariement de chaînes, qui sont généralement appliqués aux chaînes de caractères ou aux chaînes biologiques (par exemple, l'ADN ou les chaînes de protéines). Quelques-uns sont discutés dans [APO 85] et dans [CRO 94].

Généralement, les entités constituant un pattern musical n'ont pas la même prégnance : certaines notes (ou accords, etc.) ressortent davantage que d'autres en termes de position métrique, de durée, de registre, d'harmonie, de hiérarchies tonales, etc. Quel type de techniques d'appariement de formes est le plus adapté pour établir des ressemblances entre des chaînes structurées, comme des passages mélodiques ? Plus précisément, bien que l'appariement approximatif semble être la solution évidente pour saisir les variations musicales (par exemple le remplissage et « l'amincissement » du matériel thématique, des changements rythmiques, des changements de hauteur, des changements tonaux, etc.), l'appariement exact permet-il de traiter ce phénomène ?

Pour simplifier, considérons qu'il existe deux grandes approches :

- a) les techniques d'appariement approximatif de patterns (AAP) appliquées à la surface musicale non structurée ;
- b) les techniques d'appariement exact de patterns (AEP) appliquées sur la surface musicale et à diverses réductions de celle-ci en composants structurellement plus saillants.

La première approche est basée sur l'idée que les segments musicaux considérés comme parallèles (similaires) partagent *certain*s éléments identiques (par exemple, deux occurrences d'un motif mélodique auront un nombre « significatif » de notes ou d'intervalles communs, mais pas nécessairement tous). Plusieurs algorithmes d'AAP sont décrits par exemple dans [BLO 85, COP 90, ROL 96, ROW 95, STA 93]. La deuxième approche est basée sur l'idée que des segments musicaux parallèles soient nécessairement identiques du point de vue d'au moins un profil paramétrique de la surface ou d'une réduction de celle-ci (par exemple, deux occurrences d'un motif mélodique auront le même profil paramétrique au niveau superficiel ou à un certain niveau plus abstrait comme des patterns de notes/intervalles métriquement forts ou tonalement importants). On trouvera des techniques informatisées basées sur cette approche dans [CAM 98a, HIR 97].

Quels sont les avantages et les inconvénients de chacune de ces méthodologies ? Pour prendre un exemple, considérons les segments mélodiques tonaux de la figure 9.5. Quel est le degré de similitude des segments *b*, *c* et *d* avec le segment *a* ? Supposons que chaque segment mélodique est représenté sous la forme d'une séquence de notes décrites par deux valeurs : la hauteur et l'instant initial (bas de la figure 9.5).

Pour l'AAP, chacun des segments *b*, *c*, *d* est identique à 71 % au segment *a* car cinq de leurs sept notes correspondent. Selon le seuil choisi, les trois segments mélodiques ont la même similitude (et la même dissimilitude) avec *a*. Pourtant, il est

clair pour un musicien que le segment *b* est (dans la plupart des contextes tonaux) bien plus similaire à *a* que les autres car *a* et *b* se correspondent de la « bonne » façon : les notes principales sont identiques, et on ignore les « ornements », moins importantes.

a)

b) c) d)

segment a : [g, 0], [c, 4], [b, 8], [c, 9], [a, 10], [b, 11], [g, 12]
 segment b : [g, 0], [a, 2], [b, 3], [c, 4], [b, 8], [a, 10], [g, 12]
 segment c : [g, 0], [a, 4], [b, 8], [c, 9], [a, 10], [b, 11], [c, 12]
 segment d : [g, 0], [c, 4], [b, 8], [c, 9], [a, 10], [c, 11], [d, 12]

Figure 9.5. Degré de similitude des segments mélodiques *b*, *c* et *d* pour la segmentation de *a*

La deuxième méthode d'appariement de formes exige un prétraitement significatif ; par exemple, au lieu de se contenter d'examiner les segments mélodiques au niveau superficiel, on construit des niveaux abstraits de représentation qui reflètent les propriétés structurelles des segments mélodiques (par exemple des notes plus longues, des notes métriquement plus fortes ou des notes tonalement importantes). Il faut cependant souligner qu'il est possible de tenir compte dans les techniques d'AAP de la prégnance structurelle en attribuant des *poïds* aux appariements des éléments des patterns, par exemple leur contribution à la similitude pour chaque transformation, notamment relativement à la durée et aux intervalles de hauteur, comme l'ont proposé et mis en œuvre Mongeau et Sankoff [MON 90] et Rolland [ROL 99].

La seconde méthode a sur la première, l'avantage de donner des raisons explicites au jugement de similitude : les propriétés communes que l'on découvre sont codées explicitement, et cette connaissance peut être réutilisée dans d'autres tâches d'analyse ou de composition. Par contre, elle exige un traitement supplémentaire pour produire des réductions musicalement significatives de la surface.

9.4.1. Un algorithme d'induction de patterns basée sur des répétitions exactes

Un algorithme efficace permettant de calculer toutes les répétitions à l'intérieur d'une chaîne est décrit dans [CRO 81 ; Iliopoulos *et al.*, 1996]. Pour une chaîne donnée de symboles (par exemple la chaîne d'intervalles de hauteur), le processus d'appariement commence par la plus petite longueur de pattern (deux éléments) et se termine lorsqu'il a trouvé l'appariement le plus long. La complexité de cet algorithme est $O(N \cdot \log N)$, N étant la longueur de la chaîne.

Cet algorithme peut être appliqué à un nombre quelconque de profils paramétriques de la surface mélodique ou de ses réductions (par exemple : intervalles de hauteurs, contours, durées, intervalles entre instants initiaux, écarts de dynamique, etc.).

On voit que cette procédure de découverte de tous les patterns mélodiques, identiques sur un grand nombre de chaînes paramétriques mélodiques, produit un nombre très élevé de patterns possibles, dont la plupart sont peu naturels et non pertinents pour un auditeur, musicien ou analyste humain.

Une procédure a été mise au point pour attribuer une valeur de prégnance à chacun des patterns découverts, d'après les critères suivants :

- a) préférer les patterns les plus longs,
- b) préférer les patterns les plus fréquents,
- c) éviter les chevauchements.

On peut créer une *fonction de choix* qui calcule un score numérique d'un pattern selon ces principes, par exemple :

$$f(L, F, DOL) = F^a \cdot L^b / 10^c \cdot DOL$$

où L est la longueur de modèle ; F la fréquence des occurrences dans un pattern ; DOL le degré de chevauchement ; et a, b, c des paramètres auxquels on peut donner expérimentalement diverses valeurs constantes.

Chaque pattern découvert par cet algorithme reçoit un score suivant la fonction de choix.

Les patterns ayant les meilleurs scores sont considérés comme les plus significatifs (voir figure 9.6).



Figure 9.6. Frère Jacques : *principaux patterns de hauteur signalés par l'algorithme d'induction exact et la fonction de choix (appliquée aux intervalles diatoniques seulement)*

9.4.2. Un algorithme d'induction de patterns basée sur des répétitions approximatives

On l'a vu, l'induction de patterns basée sur des répétitions exactes (IPBRE) peut donner d'excellents résultats. Cependant, en plus des traitements supplémentaires requis pour la réduction de la surface musicale, il peut aussi y avoir des limitations lorsque l'on a affaire à certains genres musicaux. Le jazz improvisé en fait partie. Par exemple, dans le be-bop, genre dominant né dans les années 1940, les techniques massivement utilisées par les improvisateurs telles que l'ornementation, la variation locale et les syncopes, rendent difficile une approche par réduction-induction.

Dans le cadre de l'induction de patterns basée sur des répétitions approximatives (IPBRA), ces phénomènes locaux sont pris en charge **en autorisant en assouplissant** la contrainte d'égalité entre les différentes occurrences d'un pattern. Un *pattern* est alors défini comme un ensemble de passages (ou *segments*) mélodiques qui sont significativement similaires. Chaque segment est appelé une *occurrence* du pattern. Dans certains cas, on exige que toutes les paires de segments dans un pattern soient significativement similaires, mais nous nous concentrerons ici sur le cas où :

- l'une des occurrences (le *prototype*) est considérée comme la plus représentative du pattern ;
- chaque occurrence est significativement similaire au prototype.

Un tel *pattern en étoile* est illustré sur la figure 9.7 : le pattern apparaît non dans une seule mélodie comme dans l'exemple donné plus haut pour l'IPBRE, mais dans plusieurs (quatre mélodies).

Un certain nombre d'algorithmes d'IPBRA ont été proposés, par exemple dans Cope (1991), Rolland [ROL 99a, ROL 01b], Rowe (1993), Smaill *et al.* (1993), Stammen & Pennycook [STA 93] et Stech (1981). Comme l'expliquent Rolland & Ganascia [ROL 99b], certains d'entre eux peuvent fonctionner en temps réel, les autres étant limités au temps différé.

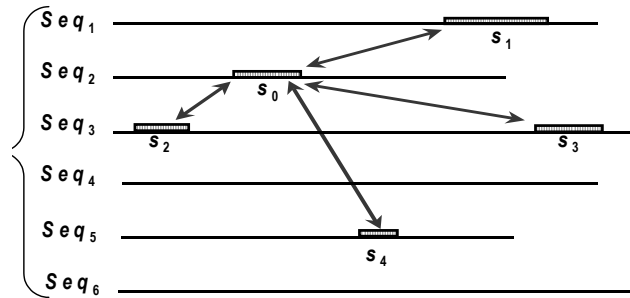


Figure 9.7. Exemple de pattern en étoile apparaissant dans quatre mélodies.
Sa liste d'occurrences est $\{s_0, s_2, \dots, s_4\}$ et le prototype est s_0

Pour formaliser la notion de « significativement similaire », on introduit une *fonction de similitude* (ou, de façon équivalente, une fonction de dissimilitude). Plus spécifiquement, une fonction de similitude normalisée *Simil* prend ses valeurs dans l'intervalle réel $[0,1]$, s_1 et s_2 étant 2 passages mélodiques, $Simil(s_1, s_2)$ est d'autant plus proche de 1 (resp. de 0) que leur similitude calculée est forte (resp. faible). Par définition, un couple de passages (s_1, s_2) est significativement similaire, ou *équipollent*, si et seulement si $Simil(s_1, s_2)$ est supérieur ou égal à un seuil prédéfini, appelé seuil d'équipollence.

Toute fonction de similitude est basée sur un *modèle de similitude* qui définit la procédure de comparaison de deux passages. C'est une caractéristique importante de tout algorithme d'IPBRA.

Dans le modèle de similitude le plus simple, on compare successivement chaque note³ du premier passage à la note correspondante du second passage et la similarité entre les deux passages est mesurée par la proportion de notes concordantes. Cette structure de correspondance fixe (« distance de Hamming ») est très limitative car elle exige que les deux passages aient le même nombre de notes. Elle est utilisée par exemple par le module d'induction de patterns d'EMI (Cope 1991).

A l'autre extrémité du spectre, figurent les modèles de similarité appartenant à la famille de la « distance d'édition ». Ils utilisent des structures de correspondance variables, afin de faire correspondre les notes des deux passages comparés de manière optimale. Une illustration en est donnée en figure 9.9 et figure 9.10, sur la

3. Les silences peuvent être traités de la même façon que les notes. Par souci de simplicité, nous ne parlons ici que des notes.

base du couple de passages de la figure 9.8. Lors de la comparaison des passages s_1 et s_2 , une note n de s_1 peut être appariée (mise en correspondance) à une note n' de s_2 . Cet appariement est appelé *remplacement-identité* si n est identique à n' , et *remplacement-substitution* si elles sont différentes. Mais n peut aussi être appariée :

- à un groupe vide (zéro note) de s_2 . Ce cas, dénommé *suppression*, s'interprète comme le fait que s_1 et s_2 seraient plus similaires en enlevant n_1 de s_1 ;
- à un groupe de plusieurs notes $\{n', n'' \dots\}$ de s_2 , n'ayant d'ailleurs pas nécessairement la même hauteur ou la même durée que n . Ce cas, la *fragmentation*, s'interprète comme le fait que s_1 et s_2 seraient plus similaires si, dans s_1 , n_1 était remplacée par le groupe $\{n', n'' \dots\}$.

De façon analogue on définit d'autres types d'appariements :

- *insertion* (zéro note de s_1 appariée avec 1 note de s_2),
- *consolidation* (l'inverse de la fragmentation),
- *suppression multiple*,
- *insertion multiple*,
- etc.

Chaque appariement est considéré comme contribuant positivement ou négativement à la similitude globale entre s_1 et s_2 . Formellement parlant, un nombre réel appelé *contribution* est attribuée à chaque appariement. A l'exception de travaux anciens, la contribution attribuée à un appariement est fonction des caractéristiques de la ou des notes sur lesquelles il porte. Par exemple, la contribution attribuée à la substitution d'une note par une note très similaire (par exemple en termes de hauteur, durée et positions métriques) est plus positive qu'une substitution par une note très différente.

Pour chaque structure de correspondance entre s_1 et s_2 (appelée *alignement*) la somme des contributions de tous les appariements donne un score de similitude. $Simil(s_1, s_2)$ est alors défini comme le plus élevé de ces scores, en d'autres termes, comme le score du meilleur alignement possible entre les deux passages.

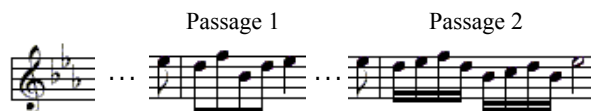


Figure 9.8. Deux passages mélodiques (J. Haydn, Concerto pour Trompette en mi bémol majeur)

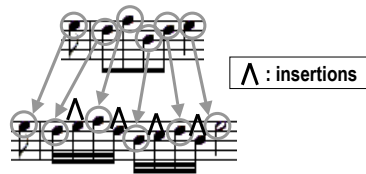


Figure 9.9. L'un des alignements possibles entre les deux passages, avec six remplacements et quatre insertions

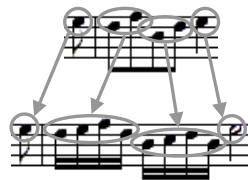


Figure 9.10. Un autre alignement entre les deux passages, avec deux remplacements individuels et deux remplacements généralisés

Le calcul de $Simil(s_1, s_2)$, prenant en compte tous les alignements possibles, est complexe et fait intervenir une combinatoire énorme. Des techniques de programmation dynamique permettent d'effectuer ce calcul en un temps proportionnel au produit des longueurs de s_1 et s_2 .

FLEXPat (*Flexible Extraction of Patterns*) est un algorithme d'IPBRA autorisant l'emploi de telles structures de correspondances variables. Une brève description de l'algorithme est donnée en annexe A.3. Pour plus de détails, on pourra se référer aux articles suivants : [ROL 99] pour une perspective musicale, [ROL 01] pour une perspective davantage formelle et technique.

FLEXPat comprend deux phases principales :

1) une phase de *construction du graphe d'équipollence* identifie toutes les paires de passages équipollents, c'est-à-dire (rappel) significativement similaires. Les calculs fortement combinatoires de cette phase sont effectués de façon économique (en termes de temps de calcul et de mémoire) grâce à l'emploi de concepts dérivés de la programmation dynamique. Une structure intermédiaire, appelée *graphe d'équipollence*, est produite dans laquelle :

- chaque sommet correspond à (pointe vers) un seul passage et un seul ;
- chaque arête traduit le fait que deux passages sont équipollents. En d'autres termes, en supposant que v correspond au passage s et que v' correspond au passage s' , l'arête (v, v') existe si et seulement si s et s' sont équipollents ;

2) une phase d'*extraction de sous-graphes* extrait les patterns proprement dits à partir du graphe d'équipollence. La forme précise des sous-graphes extraits peut être variable, par exemple :

- des sous-graphes en étoile tels que celui de la figure 9.11 ;
- des cliques maximales comme dans l'algorithme UNSCRAMBLE (voir annexe A.2).

Nous nous focaliserons ici sur le cas de sous-graphes en étoile. Dans chaque sous-graphe extrait, l'ensemble des arêtes représente un ensemble de passages qui est à son tour interprété comme l'ensemble des occurrences d'un pattern en étoile comme défini plus haut. Un prototype est ainsi obtenu pour chaque pattern extrait.

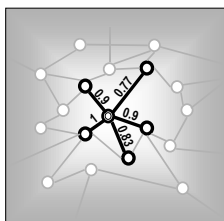


Figure 9.11. Exemple de sous-graphe en étoile dans le graphe d'équipollence

Comme pour l'algorithme d'IPBRE, un score est attribué à chaque pattern induit. Il est généralement basé sur les critères suivants :

- favoriser les patterns les plus longs,
- favoriser les patterns ayant le plus d'occurrences.

Il est à noter que, contrairement au cas « exact » où intervenait un troisième critère « éviter les recouvrements », FLEXPAT évite les recouvrements de manière plus directe : en effet, lors de la construction du graphe d'équipollence, si deux passages se recoupent trop (le seuil de recouvrement étant ajustable par l'utilisateur), leur similarité ne sera pas calculée du tout. On ne risque donc pas de les considérer comme deux occurrences d'un même pattern.

Parmi les principaux paramètres numériques de FLEXPAT, figure l'entier m_{max} qui contrôle la longueur maximum des passages constitutifs d'un pattern. En pratique, il n'y a pas grand intérêt à rechercher des patterns extrêmement longs ; $m_{max} \ll N$ où N est le nombre total de notes de la mélodie (ou de l'ensemble de mélodies) où l'on cherche à induire les patterns.

La sortie de FIEXPath est une liste de patterns en étoile classés par degrés de force décroissants. Sa complexité temporelle globale théorique est majorée par $O(m_{max}^2 \cdot N^2)$, formule que l'on peut réécrire $O(N^2)$ compte tenu de la remarque précédente. Elle est d'ailleurs bien plus faible en pratique.

Nous allons maintenant présenter un exemple de sortie de FIEXPath. L'algorithme est exécuté sur un ensemble de dix transcriptions de solos de Charlie Parker sur le schéma d'accord de Blues en do majeur : trois prises (*takes*) de *Cool Blues*, trois prises de *Relaxin' at Camarillo* et quatre prises de *Perhaps*. La longueur totale N est d'environ 2 000. Parmi les valeurs de paramètres de FIEXPath, on a fixé m_{max} à 27 et le seuil d'équipollence à 0,7 (avec une fonction de similitude normalisée)

Le graphe d'équipollence a environ 12 000 sommets et 40 000 arêtes. Le plus grand pattern en étoile a trente-quatre segments plus le prototype. Le pattern le plus long a un prototype de longueur vingt-sept (la longueur d'un pattern pouvant être définie comme la longueur de son prototype, ou comme la longueur moyenne de ses occurrences).

La figure 9.12 montre le prototype de l'un des patterns les plus longs (longueur 26) extrait par FIEXPath, que nous désignerons dans la suite par Pat1. La figure 9.13, la figure 9.14 et la figure 9.15 montrent ses trois occurrences, classées par similitude décroissante par rapport au prototype, soit 0,93, 0,76 et 0,75.

Pat1 a un sens musical. Si le prototype et les occurrences ont en commun le sous-segment final, leurs débuts sont tous différents quoique similaires. En particulier, les quatre débuts finissent par un sol dont la durée et le placement métrique sont variables d'une occurrence de Pat1 à l'autre. De plus, la mise en correspondance par le programme de deux sous-segments particuliers est intéressante : les descentes diatoniques avant le premier do dans le prototype et la première occurrence (fa-mib-ré et ré-do-si respectivement) sont appariées malgré une transposition locale. On peut aussi noter que, quoique similaires, les débuts du prototype et des occurrences n'ont pas tous la même longueur (dans l'ordre : 6, 6, 3 et 4, en excluant les silences).



Figure 9.12. Prototype du pattern Pat1 : passage [19:3-23:1] du solo sur *Relaxin' at Camarillo* – prise 4



Figure 9.13. Première occurrence de Pat1 : passage [7:4-11:1]
du solo sur Relaxin' at Camarillo – prise 4



Figure 9.14. Seconde occurrence de Pat1 : passage [32:1.75-35:1]
du solo sur Perhaps – prise 3

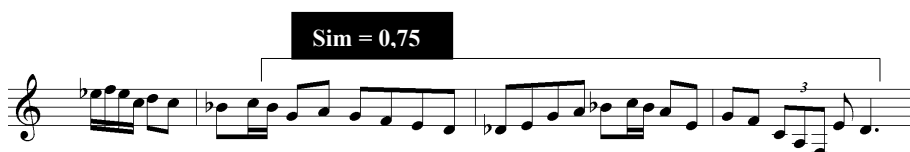


Figure 9.15. Troisième occurrence de Pat1 : passage [20:1.5-22:4]
sur Relaxin' at Camarillo – prise 3

9.5. Segmentation

La segmentation d'une surface musicale est une tâche cruciale dans l'analyse musicale : la segmentation choisie initialement peut affecter profondément la suite de l'analyse en excluant *de facto* beaucoup de structures ne correspondant pas à des segments. Nous avons dit dans la deuxième partie de ce chapitre que la segmentation d'une surface musicale est affectée non seulement par les discontinuités locales mais aussi par des processus de plus haut niveau. Le plus important de ces processus de haut niveau est peut-être le calcul de la *similitude musicale* : les patterns musicaux similaires ont tendance à être détectés et perçus comme des unités dont le début et la fin de points influencent la segmentation de la surface musicale.

Par exemple, un certain modèle de détermination des frontières locales pourrait situer une frontière locale à l'intervalle entre la troisième et la quatrième note de *Frère Jacques* comme sur la figure 9.6 (puisqu'il s'agit d'un intervalle de hauteur plus

grand que ses voisins), alors qu'il est évident qu'il y a une frontière légitime entre la quatrième et la cinquième note du fait de la répétition mélodique.

Les procédures de segmentation basées sur la similitude sont très rares (voir [BAK 89, MAR 01]). Dans cette section, nous présentons une procédure basée sur l'algorithme d'induction de patterns décrit au paragraphe 9.4.1.

L'approche basée sur l'algorithme d'appariement exact que nous avons mentionné et sur la fonction de choix permet de découvrir des patterns mélodiques « significatifs », mais il faut encore appliquer un nouveau traitement qui produira une « bonne » description de la surface (en termes d'exhaustivité, d'économie, de simplicité, etc.). Quelques-uns des patterns de hauteur devront sans doute être éliminés, et il reste possible qu'une combinaison des patterns ayant un score non maximal donne une meilleure description de la surface musicale (par exemple, dans la figure 9.6, on peut se demander s'il faut préférer, comme pattern de hauteurs, a , b , ou une combinaison des deux comme $a-a-b-b$).

Pour résoudre ce problème, on a inventé une méthodologie très simple à utiliser, mais fruste, consistant à appliquer la procédure d'induction de patterns à autant de chaînes paramétriques que nécessaire de la surface mélodique et de ses réductions. Aucun pattern n'est ignoré ; chacun contribue à chaque frontière possible de la séquence mélodique proportionnellement à son score selon la fonction de choix. Pour chaque point de la surface mélodique, on ajoute le score de choix de tous les patterns dont un des arcs le touche, de façon à créer un profil de score de frontières de patterns, normalisé à l'intervalle $[0,1]$. On suppose que les points de la surface auxquels correspondent des maximums locaux sont plus susceptibles d'être perçus comme des frontières à cause de la similitude musicale. Dans l'exemple mélodique de la figure 9.16, le profil de relief de frontières de patterns a été calculé en appliquant l'algorithme d'induction exacte de pattern aux profils paramétriques d'intervalles diatoniques, de type d'intervalles et de durée. On peut observer dans la figure les fortes limites de patterns aux points indiqués par des astérisques, où les frontières locales sont très faibles.

Les frontières découvertes par l'induction de patterns peuvent être complémentaires des frontières locales détectées avec le *MDFL* pour la définition du profil de frontière globale (PFG). Le PFG est obtenu en calculant la moyenne pondérée des profils de frontières locales (PFL) et des profils de frontières de patterns (PFP), normalisés eux aussi. Dans l'application dont il est question ici, les deux profils de force ont une importance égale. Les maximums locaux du profil de frontière globale peuvent guider la segmentation de la surface musicale.



Figure 9.16. Profil de frontières locales (PFL), profil de frontières de patterns (PFP) et profil de frontière globale (PFG) du premier thème de la Valse opus 18 de Chopin

On peut choisir une ou plusieurs segmentations dans le PFG. Dans cet exemple, nous choisissons une segmentation constituée de tous les maximums locaux de valeur supérieure à 0,7 (voir la segmentation A dans la figure 9.17). Nous utilisons aussi une autre segmentation métriquement régulière basée sur le fait que la structure métrique et la structure de groupement tendent à coïncider (en phase et hors phase) ; ce groupement basé sur la coïncidence peut être déterminé en trouvant simplement la meilleure adéquation du réseau métrique précédemment découvert (3/4 dans notre exemple) sur le PFG (segmentation B, figure 9.17).

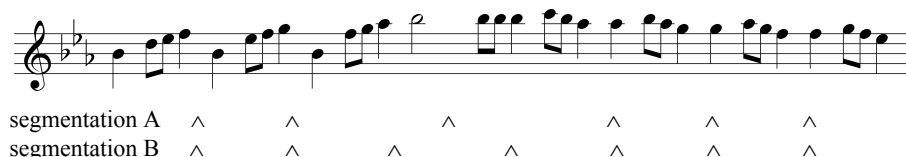


Figure 9.17. Deux segmentations de la mélodie initiale de la Valse de Chopin opus 18. La segmentation A provient du PFG de la figure 9.16 en retenant les maximums locaux $\geq 0,7$. La segmentation B est superposable et en phase avec la structure métrique en 3/4 de la mélodie.

Nous allons maintenant voir comment organiser en catégories utiles (motifs, thèmes, etc.) les segments mélodiques produits lors de cette segmentation.

9.6. Similitude musicale et catégorisation

La similitude musicale semble être un phénomène difficile à décrire de façon systématique. Quand deux passages musicaux sont-ils similaires, et quand sont-ils suffisamment différents pour être considérés comme dissemblables ? Quels passages musicaux sont assez similaires pour appartenir à la même catégorie ? Comment traiter les passages ambigus ?

Höthker *et al.*, [HOT 00] et Rolland [ROL 99] mentionnent quelques tentatives récentes d'application de modèles informatiques de catégorisation aux passages mélodiques. Tous les modèles informatiques de la catégorisation musicale doivent répondre aux problèmes de la détermination du nombre optimum de catégories musicales, de la détermination d'un seuil de similitude et du traitement des segments mélodiques ambigus. Nous proposons ci-dessous une façon de répondre à cette question.

Nous pensons qu'il existe une forte corrélation entre les notions de catégorisation, de similitude et de représentation des entités ou des propriétés. Les gens ne partent pas d'une description précise d'entités et de propriétés pour trouver des similarités entre eux puis des regroupements en catégories constituées selon ces similarités (figure 9.18a). Il est plus vraisemblable qu'en organisant leur connaissance du monde, ils modifient leurs représentations d'entités en même temps qu'ils créent des catégories (émergentes) et portent des jugements de similitude (figure 9.18b).

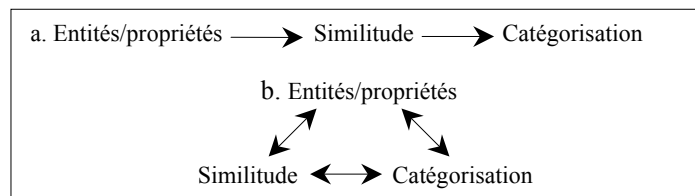



Figure 9.18. Relations entre entités/propriétés, similitude et catégorisation

Une de nos principales hypothèses est que la similitude dépend toujours du contexte (elle est contextuelle) et que les cas où elle semble relativement stable sont simplement les cas dans lesquels le contexte (par exemple la structure du monde ou un système culturel) est lui-même stable. Certes, il existe des contraintes perceptives générales quant à ce qui est perceptible, mais au-delà de ces contraintes, les propriétés des entités acquièrent leur prégnance dans un contexte donné, pour une certaine tâche de catégorisation ou pour certain un jugement de similitude. **Tversky (1977)** a montré l'importance du contexte dans les jugements de similitude et la façon dont les propriétés d'objets prennent une valeur *diagnostique* dans un contexte spécifique ; cependant il considère ces effets contextuels sur la similitude comme

des exceptions et non comme la norme, et définit la similitude indépendamment de la catégorisation.

Dans le domaine musical, fortement autoréférentiel, on peut considérer une œuvre (ou un ensemble d'œuvres) comme un contexte local comportant des motifs, des thèmes, des groupes de progression harmoniques, etc. La recherche de la similitude de deux passages musicaux isolés a généralement des résultats douteux ou peu intéressants. Par exemple, les deux patterns rythmiques  sont-ils similaires ? Certaines personnes diront que oui, d'autre que non, et d'autres ne se prononceront pas. Nous pensons quant à nous que la question est mal posée et que l'on a tout lieu de ne pas se prononcer. Le problème est que ces deux patterns sont privés de contexte. Par exemple, on peut les considérer comme très différents dans un contexte musical constitué seulement de ces deux patterns (par exemple le final de la *Sonate pour violon en do majeur BWV 1005* de J.S. Bach), mais très semblables dans un contexte plus divers contenant aussi des triplets et d'autres patterns rythmiques (par exemple la première section de la *Sonate en fa majeur KV 280* de Mozart). Le contexte semble donc primordial lorsque nous faisons des jugements de ressemblance et que nous créons des catégories concernant les passages musicaux, et il n'y a pas de critère absolu servant à dire *de façon générale* quels objets sont similaires.

Sur ces principes, nous avons mis au point l'algorithme *Unscramble* [CAM 00 ; CAM 01b] qui, à partir d'un ensemble d'objets et d'un ensemble initial de propriétés, produit une série de groupes plausibles pour un contexte donné. Au cours de ce processus dynamique, l'ensemble de propriétés est ajusté pour permettre d'obtenir une description satisfaisante. Cet algorithme de regroupement détermine automatiquement un nombre approprié de groupes, choisit les attributs caractéristiques ou définitoires de chaque catégorie et permet un chevauchement limité des groupes. L'algorithme est brièvement décrit en annexe A.1. Pour une description plus détaillée, on pourra se reporter à [CAM 00].

Cet algorithme de regroupement est appliqué aux segments mélodiques illustrés dans la figure 9.17, de façon à calculer au niveau superficiel, pour chaque segment mélodique, les attributs suivants :

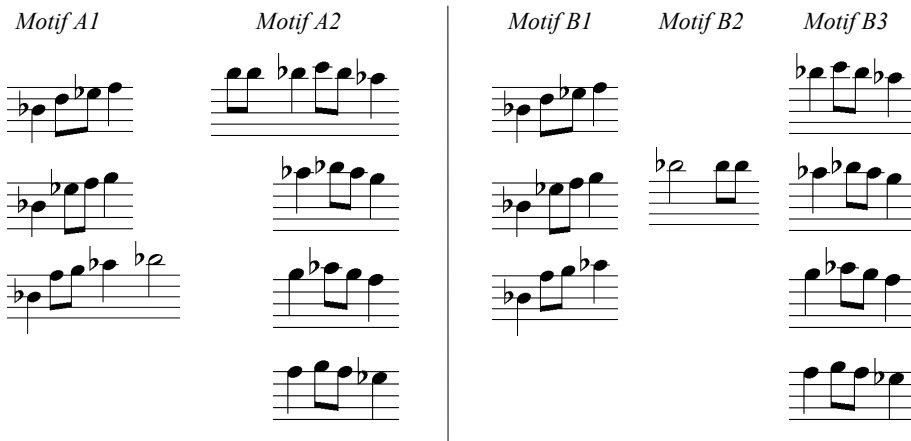
- divers attributs d'intervalle de hauteur, comme les patterns d'intervalles mesurés en demi-tons, les intervalles diatoniques, le type d'intervalle mélodique⁴ et le contour (ascendant, *recto tono*, descendant) ;

4. Type d'intervalle mélodique : (E)gal : 0 demi-tons, (C)onjoint : 1 ou 2 demi-tons, (D)isjoint : ≥ 3 demi-tons.

– divers attributs rythmiques, comme les patterns d’intervalles entre instants initiaux, les rapports de durée⁵ et les comparaisons de durée de séquences (PC : plus courte, PL : plus longue, E : égale).

Par exemple, dans le premier segment de la figure 9.17a, les attributs, dans l’ordre, ont les valeurs suivantes :

- attributs d’intervalles de hauteur : 4-1-2, 2-1-1, PL-PC-PC, A-A-A ;
- attributs rythmiques : 1/4-1/8-1/8-1/4, 1/2-0-2, PC-E-PL.



a) Regroupements pour la segmentation A b) Regroupements pour la segmentation B

Figure 9.19. Motifs produits par l’algorithme de regroupement pour les deux segmentations (figure 9.17) de la mélodie initiale de la Valse opus 18 de Chopin (les colonnes de segments mélodiques représentent les groupes)

L’algorithme *Unscramble* est alors appliqué aux sept segments de la segmentation A, puis aux huit segments de la segmentation B. Les résultats du processus de regroupement sont indiqués dans la figure 9.19. Soulignons que des modèles d’attributs sont appariés s’ils sont identiques ou si l’un d’eux est un sous-modèle contigu exact de l’autre (par exemple PL-PC-PC est un sous-modèle de PL-PC-PC-PC). Non seulement l’algorithme produit des regroupements de segments mais il choisit aussi les ensembles d’attributs les plus caractéristiques de chaque groupe. Par exemple, le motif B₁ est défini par les attributs d’intervalles de degrés et de contour

5. Rapport de durée : proportion entre deux intervalles successifs, par exemple 1/2 pour une séquence noire-croche.

qui reçoivent un poids maximum de 1 tandis que les attributs rythmiques reçoivent des valeurs très faibles (tous les passages correspondant au motif B_1 ont le même pattern rythmique, mais ce pattern est aussi celui de tous les passages correspondant de B_3 ; ce pattern rythmique n'est donc pas caractéristique de B_1).

9.7. Conclusion

Dans ce chapitre, nous avons présenté un certain nombre de techniques informatiques et d'algorithmes pour essayer de décrire la structure musicale. La possibilité de mettre en œuvre des programmes capables de repérer les aspects de base de la « compréhension » musicale est cruciale, non seulement pour améliorer notre compréhension de la musique elle-même, mais aussi pour construire les outils analytiques qui peuvent être utilisés dans un grand nombre d'applications musicales pratiques, comme l'exécution interactive, la composition assistée par ordinateur, l'indexation et la recherche de données musicale, l'expressivité de l'exécution automatique, etc.

Dans ce chapitre, nous nous sommes concentrés sur des méthodes d'analyse de mélodies en termes de frontières locales, de segmentation, de patterns, de motifs, etc., mais ces techniques, étant basées sur des principes logiques et cognitifs généraux, peuvent être adaptées à d'autres sortes de tâches analytiques et concrétisées dans des programmes musicaux plus sophistiqués. Le principal objectif du chapitre était d'illustrer divers types de problèmes du traitement de la musique et quelques façons de les résoudre, plutôt que de proposer des programmes prêts à l'emploi. L'analyse musicale automatisée est un vaste territoire qui reste en grande partie inexploré.

9.8. Bibliographie

- [APO 85] APOSTOLICO A., GALIL Z. (DIR.), *Combinatorial Algorithms on Words*, Série NATO ASI, Springer-Verlag, Berlin, 1985.
- [BAK 89] BAKER M., « An Artificial Intelligence Approach to Musical Grouping Analysis », *Contemporary Music Review*, 3 :43-68, 1989.
- [BAT 98] BATTEL G.U., FIMBIANTI R., « Aesthetic Quality of Statistic Average Music Performance in Different Expressive Intentions », *Proceedings of the XII Colloquium of Musical Informatics*, Gorizia, Italie, 1998.
- [BEN 80] BENT I.D., « Analysis », dans S. Sadie (dir.), *The New Grove Dictionary of Music and Musicians*, vol. 1, Macmillan, Londres, 1980.
- [BLO 85] BLOCH J., DANNENBERG R., « Real-Time Accompaniment of Polyphonic Keyboard Performance », *Proceedings of the 1985 International Computer Music Conference*, p. 279-290, 1985.

- [CAM 96] CAMBOUROPOULOS E., « A Formal Theory for the Discovery of Local Boundaries in a Melodic Surface », *Proceedings of the III Journées d' Informatique Musicale*, Caen, France, 1996.
- [CAM 97] CAMBOUROPOULOS E., « Musical Rhythm : Inferring Accentuation and Metrical Structure from Grouping Structure », dans M. Leman (dir.), *Music, Gestalt and Computing – Studies in Systematic and Cognitive Musicology*, Springer-Verlag, Berlin, 1997.
- [CAM 98a] CAMBOUROPOULOS E., « Musical Parallelism and Melodic Segmentation », *Proceedings of the XII Colloquium of Musical Informatics*, Gorizia, Italie, 1998.
- [CAM 98b] CAMBOUROPOULOS E., *Towards a General Computational Theory of Musical Structure*, Thèse de doctorat, Faculté de Musique et département de l'intelligence artificielle, Université d'Edimbourg, 1998.
- [CAM 00] CAMBOUROPOULOS E., Widmer G., « Automated Motivic Analysis Via Melodic Clustering », *Journal of New Music Research*, 29(4) : 303-318, 2000.
- [CAM 01a] CAMBOUROPOULOS E., « The *Local Boundary Detection Model (LBDM)* and its application in the study of expressive timing », *Proceedings of the International Computer Music Conference (ICMC01)*, Havana, Cuba, 2001.
- [CAM 01b] CAMBOUROPOULOS E., « Melodic Cue Abstraction, Similarity and Category Formation : A Formal Model », *Music Perception*, 18(3) :347-370 2001.
- [CAM 01] CAMBOUROPOULOS E., CRAWFORD T., ILIOPOULOS C.S., « Pattern Processing in Melodic Sequences : Challenges, Caveats and Prospects », *Computers and the Humanities*, 35(1) :9-21, 2001.
- [CLA 85] CLARKE E.F., « Structure and Expression in Rhythmic Performance », dans P. Howell *et al.* (dir.), *Musical Structure and Cognition*, Academic Press, Londres, 1985.
- [CON 95] CONKLIN D., WITTEN I.H., « Multiple Viewpoint Systems for Music Prediction », *Journal of New Music Research*, 24 :51-73, 1995.
- [COP 90] COPE D., « Pattern-Matching as an Engine for the Computer Simulation of Musical Style », *Proceedings of the International Computer Music Conference*, Glasgow, 1990.
- [CRA 98] CRAWFORD T., ILIOPOULOS C.S., RAMAN R., « String Matching Techniques for Musical Similarity and Melodic Recognition », *Computing in Musicology*, 11 :71-100, 1998.
- [CRO 81] CROCHEMORE M., « An Optimal Algorithm for Computing the Repetitions in a Word », *Information Processing Letters*, 12(5) :244-250, 1981.
- [CRO 94] CROCHEMORE M., RYTTER W., *Text Algorithms*, Oxford University Press, Oxford, 1994.
- [DEL 87] DELIÈGE I., « Grouping Conditions in Listening to Music : An Approach to Lerdahl and Jackendoff's Grouping Preference Rules », *Music Perception*, 4 :325-360, 1987.
- [EPS 95] EPSTEIN D., *Shaping Time : Music, the Mind and Performance*, Schirmer books, New York, 1995.

- [HIR 97] HIRAGA Y., « Structural Recognition of Music by Pattern Matching », *Proceedings of the International Computer Music Conference*, Thessaloniki, Grèce, 1997
- [HOT 00] HÖTHKER K., HÖRNEL D., ANAGNOSTOPOULOU C., « Investigating the Influence of Representations and Algorithms in Music Classification », *Computing in the Humanities*, 35(1) :65-79, 2000.
- [HOT 02] HÖTHKER K., THOM B., SPEVAK C., « Melodic Segmentation : Evaluating the Performance of Algorithms and Musical Experts », *Proceedings of the International Conference on Music and AI (ICMAI02)*, Edinbourg, (publié par Springer-Verlag, Berlin, Heidelberg), 2002.
- [JAC 87] JACKENDOFF R., *Consciousness and the Computational Mind*, The MIT Press, Cambridge, 1987.
- [LEE 91] LEE C.S., « The Perception of Metrical Structure : Experimental Evidence and a Model », dans P. Howell *et al.* (dir.), *Representing Musical Structure*, Academic Press, Londres, 1991.
- [LER 83] LERDAHL F., JACKENDOFF R., *A generative Theory of Tonal Music*, The MIT Press, Cambridge, 1983.
- [LON 82] LONGUET-HIGGINS H.C., LEE C.S., « The Perception of Musical Rhythms », *Perception*, 11 :115-128, 1982.
- [MAR 01] MARSDEN A., « Representing Melodic Patterns as Networks of Elaborations », *Computers and the Humanities*, 35 :37-54, 2001.
- [MEL 00] MELUCCI M., ORIO N., « A Novel Methodology for Music Information Retrieval », *Proceedings of the XIII Colloquium on Musical Informatics*, L'Aquila, Italie, 2000.
- [MON 90] MONGEAU M., SANKOFF D., « Comparison of Musical Sequences », *Computer and the Humanities*, 24 :161-175, 1990.
- [PAR 94] PARNCUTT R., « Template-Matching Models of Musical Pitch and Rhythm Perception », *Journal of New Music Research*, 23 :145-167, 1994.
- [POV 85] POVEL D.J., ESSENS P., « Perception of Temporal Patterns », *Music Perception*, 2 :411-440, 1985.
- [ROL 96] ROLLAND P.Y., GANASCIA J.G., « Automated Extraction of Prominent Motives in Jazz Solo Corpuses », *Proceedings of the 4th International Conference on Music Perception and Cognition (ICMPC'96)*, p. 491-495, Montréal, août 1996.
- [ROL 99a] ROLLAND P.Y., « Discovering Patterns in Musical Sequences », *Journal of New Music Research*, 28 :4, p. 334-350, décembre 1999.
- [ROL 99b] ROLLAND P.Y., GANASCIA J.G., « Musical Pattern Extraction and Similarity Assessment », dans E. Miranda (dir.), *Readings in Music and Artificial Intelligence*, Gordon & Breach, New York, 1999.
- [ROL 01] ROLLAND P.Y., « FIEXPAT: Flexible Extraction of Sequential Patterns », *Proceedings IEEE International Conference on Data Mining (IEEE ICDM'01)*, San Jose, Californie, 29 novembre- 2 décembre, 2001.

- [ROS 92] ROSENTHAL D., « Emulation of Human Rhythm Perception », *Computer Music Journal*, 16(10) :64-76, 1992.
- [ROW 95] ROWE R., LI T.C., « Pattern Processing in Music », *Proceedings of the Fifth Biennial Symposium for Arts and Technology*, Connecticut College, New London, 1995.
- [STA 93] STAMMEN D.R., PENNYCOOK B., « Real-time Recognition of Melodic Fragments Using the Dynamic Timewarp Algorithm », *Proceedings of the International Computer Music Conference (ICMC'93)*, 1993.
- [STE 77] STEEDMAN M.J., « The Perception of Musical Rhythm and Metre », *Perception*, 6 :555-569, 1977.
- [TEM 01] TEMPERLEY D., *The Cognition of Musical Structures*, The MIT Press, Cambridge, 2001.
- [TEN 80] TENNEY J., POLANSKY L., « Temporal Gestalt Perception in Music », *Journal of Music Theory*, 24 : 205-241, 1980.