N° d'ordre : 4406

THÈSE

PRÉSENTÉE À

L'UNIVERSITÉ BORDEAUX I

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET D'INFORMATIQUE

Par **Thomas Rocher**

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : INFORMATIQUE

ANALYSE ET MODÉLISATION DES INFORMATIONS TONALES DANS LA MUSIQUE OCCIDENTALE

Soutenue le :

Après avis des rapporteurs :

François Pachet	Chercheur
Gaël Richard	Directeur de Recherche

Devant la commission d'examen composée de :

Darrell Conklin	Professeur	Examinateur
Myriam Desainte-Catherine	Professeur	Examinateur
Yves Grenier	Professeur	Président
Pierre Hanna	Maître de Conférences	Examinateur
François Pachet	Chercheur	Rapporteur
Gaël Richard	Directeur de Recherche	Rapporteur
Matthias Robine	Maître de Conférences	Examinateur

– 2011 –

Résumé

Les travaux présentés dans cette thèse s'inscrivent dans le cadre de l'informatique musicale, et plus particulièrement dans le domaine de la modélisation et de l'analyse de l'harmonie.
Nous présentons tout d'abord cinq informations tonales distinctes : les notes de musique, les groupes de notes (ensemble des notes de musique jouées simultanément), les accords, les tonalités locales et la tonalité générale d'un morceau. Nous proposons ensuite de les organiser dans un nouveau modèle hiérarchique de l'harmonie, appelé arbre d'harmonie.
Les méthodes d'estimation de ces informations tonales sont ensuite abordées. Nous proposons d'abord certaines améliorations pour l'estimation des accords et de la tonalité générale depuis une représentation symbolique de la musique, avant de présenter une nouvelle méthode d'estimation simultanée des accords et des tonalités locales d'un morceau de musique. Le système proposé pour une telle estimation est utilisable depuis une représentation symbolique de la musique comme depuis le signal audio. L'évaluation de cette nouvelle méthode met en évidence le fait que l'estimation simultanée profite à la précision de l'estimation des accords comme à celle des tonalités, et qu'elle permet d'atteindre une précision supérieure à celle d'une des méthodes de l'état de l'art les plus performantes.
Nous présentons pour finir l'application de la présente étude à la recherche de similarité musicale, en démontrant, à l'aide d'exemples de plagiat, l'efficacité du nouveau modèle de l'harmonie proposé pour la comparaison musicale.

Abstract

The work presented in this thesis belongs to the field of computer music analysis, and especially to harmony modelling and analysis.
We first present five kinds of tonal properties : music notes, simultaneities, chords, local keys and keys. A new hierarchical model of harmony is then proposed, and named harmony tree. Estimation of those tonal properties are then studied. Improvements of state-of-the-art methods are proposed regarding key and chord estimation from a symbolic representation of music. We then propose a new method for concurrent estimation of chord and local key, both from a symbolic representation and from audio. Evaluations show that estimating chords and keys altogether benefits the chord estimation as well as the local key estimation. The accuracy of the proposed methods outperforms some of the best sate-of-the-art methods when it comes to chord recognition.
Finally, the contribution of the proposed model of harmony to the musical similarity estimation is presented. Results show that music comparison using this new model offers a better precision by taking concurrently into account different levels of the harmony.

Table des matières

1 Introduction **3**

1.1 Description des paramètres musicaux 5

1.1.1 Rythme 5

1.1.2 Timbre 6

1.1.3 Forme 7

1.1.4 Mélodie 7

1.1.5 Harmonie 7

1.2 Résumés des chapitres 8

2 Les informations tonales **11**

2.1 Définitions 11

2.1.1 Note de musique 12

2.1.1.1 Intervalle 12

2.1.1.2 Note de passage et ornement 13

2.1.1.3 Notation anglo-saxonne 13

2.1.1.4 Paramétrisation 14

2.1.2 Groupe de notes 14

2.1.2.1 Chiffrage 15

2.1.2.2 Paramétrisation 15

2.1.3 Accord 16

2.1.3.1 Paramétrisation 17

2.1.3.2 Différence entre *groupe de notes* et *accord* 18

2.1.4 Tonalité locale 18

2.1.4.1 Paramétrisation 19

2.1.5 Tonalité principale 19

2.1.5.1 Gamme et mode 20

2.1.5.2 Cycle des quintes 20

2.1.5.3 Paramétrisation 20

2.2 Descripteurs des informations tonales 23

2.2.1 Depuis une représentation symbolique de la musique 23
2.2.1.1 Description 23
2.2.1.2 Profils de notes 24
2.2.2 Depuis le signal audio 25
2.2.2.1 Description 25
2.2.2.2 Vecteurs chroma 29
2.3 Conclusion 30

3 Modèle de l'harmonie 33
3.1 Distances entre informations tonales 33
3.1.1 Distances entre notes dans l'harmonie 33
3.1.2 Profils de tonalité 34
3.1.3 Mesure entre accords dans l'harmonie 35
3.1.3.1 Affinités entre accords 36
3.1.3.2 Contenu d'intervalle 37
3.1.3.3 Distance d'Estrada 38
3.1.3.4 Distances géométriques 38
3.1.3.5 Espace tonal (*Tonal Pitch Space*) 40
3.1.3.6 Distance de Paiement 41
3.1.3.7 Distance de *Pitch-Interval* 42
3.1.4 Propriétés des différentes distances et exemples 43
3.2 Représentations hiérarchiques 43
3.2.1 Notions sur les arbres 43
3.2.2 Analyse schenkerienne 45
3.2.3 Théorie Génératrice de la Musique Tonale 46
3.2.4 Le modèle cognitif de Bharucha 48
3.2.5 Le modèle hiérarchique de Rizo 51
3.3 Arbre d'harmonie 52
3.3.1 Description 52
3.3.2 Propriétés 56

4 Analyse des informations tonales 57
4.1 Notes et groupes de notes 57
4.1.1 Analyse depuis une représentation symbolique 57
4.1.2 Analyse depuis le signal audio 59
4.1.2.1 Transcription monophonique 60
4.1.2.2 Transcription polyphonique 60
4.2 Accords 62
4.2.1 Depuis une représentation symbolique de la musique 62

4.2.1.1 Méthodes existantes 63
4.2.1.2 Méthode directe : *MD* 64
4.2.1.3 Méthode utilisant les groupes de notes : *MGN* 64
4.2.1.4 Résultats 69
4.2.1.5 Conclusion 70
4.2.2 Depuis le signal audio 71
4.2.2.1 Recherche de modèles d'accords 71
4.2.2.2 Modèles statistiques 72
4.2.2.3 Modèles musicaux / hybrides 74
4.3 Tonalités 76
4.3.1 Depuis une représentation symbolique de la musique 76
4.3.1.1 Une première proposition d'amélioration pour les méthodes d'analyse de tonalité 79
4.3.1.2 Utilisation des accords pour l'analyse de tonalité 83
4.3.2 Depuis le signal audio 88
4.3.2.1 Adaptations de l'algorithme KS 88
4.3.2.2 Méthodes utilisant une approche statistique 89

5 Estimation conjointe des accords et des tonalités 91
5.1 Principe général 91
5.1.1 Segmentation de l'entrée musicale 92
5.1.2 Énumération de *couples candidats* 93
5.1.3 Création d'un graphe de couples candidats 93
5.1.4 Sélection du chemin optimal 95
5.2 Application depuis une représentation symbolique de la musique 97
5.2.1 Segmentation 97
5.2.2 Énumération des candidats 97
5.2.2.1 Tonalités locales candidates 97
5.2.2.2 Accords candidats 98
5.2.3 Évaluation et résultats 99
5.2.3.1 Distances musicales 99
5.2.3.2 Évaluation de l'énumération des candidats 103
5.2.3.3 Comparaison à *Melisma* 105
5.2.4 Conclusion 108
5.3 Application depuis le signal audio 108
5.3.1 Calcul de vecteurs chroma 108
5.3.2 Énumérations des candidats 110
5.3.2.1 Vecteurs chroma et approche multi-résolutions 110
5.3.2.2 Vecteurs modèles d'accords et de tonalités 111

5.3.3 Évaluation . . . 111
5.3.3.1 Paramètres des vecteurs chroma . . . 113
5.3.3.2 Nombre et origine des candidats . . . 115
5.3.3.3 Estimation des tonalités locales . . . 117
5.3.3.4 Lissage final . . . 118
5.3.3.5 Comparaison à une méthode de l'état de l'art pour les accords. 119
5.3.3.6 Comparaison à une méthode directe pour les tonalités locales 119
5.3.3.7 Influence de l'estimation simultanée . . . 120
5.3.4 Conclusion . . . 120
5.4 Conclusion . . . 121

6 Estimation de similarité harmonique 123
6.1 Similarité entre séquences d'accords . . . 124
6.1.1 Alignement local . . . 125
6.1.2 Représentation des séquences d'accords . . . 126
6.1.3 Évaluation et résultats . . . 128
6.1.3.1 Bases de données . . . 128
6.1.3.2 Protocole d'évaluation . . . 129
6.1.3.3 Résultats . . . 130
6.1.3.4 Conclusion . . . 131
6.2 Similarité harmonique à partir de séquences d'accord estimées . . . 132
6.2.1 Similarité harmonique de musique symbolique . . . 132
6.2.2 Similarité harmonique de signaux sonores musicaux . . . 133
6.3 Utilisation du modèle en arbre de l'harmonie . . . 135
6.3.1 Exemple de plagiat . . . 135
6.3.2 Expériences sur 4 cas de plagiat . . . 136
6.3.3 Conclusion . . . 138

7 Perspectives et conclusion 141
7.1 Perspectives . . . 141
7.1.1 Temps réel . . . 141
7.1.2 Dictionnaire d'accords étendu . . . 142
7.1.3 Intégration des informations de structure . . . 144
7.1.4 Utilisation de regroupements successifs de vecteurs chroma . . . 146
7.2 Conclusion . . . 149

Bibliographie 154

A Annexes : bases de données **167**
A.1 Fichiers MIDI pour analyse des accords 167
A.1.1 12 morceaux pop 167
A.1.2 148 morceaux d'accompagnement 167
A.1.3 57 morceaux d'accompagnement pour l'évaluation des distances . . . 173
A.2 Fichiers MIDI pour analyse de tonalité 173
A.2.1 Extraits de la base d'apprentissage du MIREX 173
A.2.2 Chansons traditionnelles finnoises 174
A.2.3 Extraits de la discographie des Beatles 174
A.3 Enregistrements studio des Beatles 176

Liste des tableaux

1 Noms français et noms anglo-saxon des notes de musique. 13
2 Les 12 classes de notes représentant les 12 dimensions d'un profil de notes. . 24
3 Correspondance entre différentes notes et leurs fréquences fondamentales associées. 29
4 Les profils de tonalités de Krumhansl et Temperley. 36
5 Vecteur de potentiel d'affinité pour la classe de Do. 36
6 Vecteur de potentiel d'affinité pour l'accord de Do Majeur. 36
7 Valeurs d'instabilités normalisés pour la tonalité de Do Majeur et Do mineur 42
8 Valeurs d'instabilités de chaque intervalle. 42
9 Mesures d'enchaînements d'accords selon différentes distances 44
10 Résultats de précision des méthodes d'estimation d'accords MD et MGN. . 69
11 Déviations de l'accordage sur les enregistrements des Beatles. 72
12 Précision de l'estimation de tonalité selon la méthode étudiée. 81
13 Proposition de correction de la méthode hybride pour l'estimation de tonalité. 82
14 Influence de l'origine des notes sur l'estimation de tonalité. 87
15 Résultats de l'estimation d'accords selon la distance musicale utilisée comme calcul de coût dans le système proposé. 101
16 Test de différence statistique significative des distances musicales. 102
17 Pourcentage d'efficacité des différentes configurations de règles pour l'estimation des fondamentales des accords. 105
18 Évaluation de l'estimation des fondamentales des accords. 106
19 Évaluation de l'estimation des fondamentales et des modes des accords. . . . 107
20 Potentiel énuméré en fonction du nombre de candidats accords énumérés, pour deux dimensions de vecteurs chroma différentes. 113
21 Potentiel énuméré en fonction du nombre de candidats accords énumérés, pour différentes tailles de trames différentes. 114
22 Potentiel énuméré en fonction du nombre de candidats accords énumérés, avec et sans filtre médian. 115

23 Potentiel énuméré et précision du système, en fonction du nombre de candidats accords énumérés. 116
24 Potentiel énuméré, nombre d'accords candidats distincts énumérés, et précision du système pour les configurations bi-chroma et tri-chroma. 117
25 Analyse de l'erreur sur l'estimation de tonalités locales. 118
26 Comparaison avec une méthode de l'état de l'art pour l'estimation des accords. 119
27 Analyse de l'erreur pour l'estimation de tonalités locales. 120
28 Illustration du bénéfice de l'estimation simultanée accord/tonalité locale. . . 120
29 Score de substitution entre deux accords a_1 et a_2 en fonction de l'intervalle entre leurs fondamentales. 127
30 Trois versions différentes d'un même titre. 128
31 Efficacité des différentes représentations des séquences d'accords pour le calcul de similarité sur la base *Jazz*. 129
32 Efficacité des différents scores de substitution pour le calcul de similarité sur la base *Jazz*. 131
33 Résultats du calcul de similarité sur la base *Mix*. 131
34 Résultats en utilisant le système d'estimation d'accords. 133
35 Résultats du MIREX 2010 pour la tâche *Audio Cover Song Identification* . 134
36 Résultats de l'estimation de similarité sur quatre cas de plagiat avérés. . . . 139
37 Potentiel énuméré, nombre d'accords candidats distincts énumérés et précision du système pour les configurations tenant compte des répétitions des vecteurs chroma. 146

Table des figures

1 Une représentation des accords (au-dessus des paroles) pour la chanson *Heartbreaker* de Pat Benatar. . . . 4
2 Image représentative du domaine de recherche MIR (*Music Information Retrieval*), d'après [Fin04] . . . 6
3 Exemple de mélodie en La Majeur. . . . 7
4 Quatre variations sur *Day Tripper* des Beatles. . . . 12
5 Les intervalles musicaux. . . . 13
6 Illustration des notes de passage. . . . 13
7 Illustration de la transformation homorythmique. . . . 14
8 Chiffrage de groupes de notes. . . . 15
9 Équivalence entre groupes de notes. . . . 16
10 Une représentation courante des accords (au-dessus des paroles) pour la chanson *Hotel California* du groupe The Eagles. . . . 17
11 Le nom des notes dans l'accord. . . . 17
12 Un extrait des différents accords de Do existants. . . . 18
13 Exemple illustrant la différence entre un groupe de notes et un accord. . . . 19
14 Les trois gammes de Do correspondant aux trois modes les plus couramment rencontrés. . . . 21
15 Les différentes tonalités organisées autour du cycle des quintes. . . . 22
16 Récapitulatif des informations tonales et de leurs paramètres. . . . 23
17 Exemple de différents événements MIDI. . . . 24
18 Représentation en *Piano Roll*. . . . 24
19 Premières mesures du menuet BWV-114 de J.S. Bach. . . . 25
20 Les profils de notes respectifs des 4 premières mesures du menuet BWV-114 de J.S. Bach, suivi du profil de notes sur l'extrait entier. . . . 26
21 Illustration de l'échantillonnage et de la quantification de l'onde sonore, dans un processus de numérisation. . . . 27
22 Un signal audio représenté en amplitude en fonction du temps (en haut) et en fréquences en fonction du temps (en bas). . . . 28

23 Les 10 premières harmoniques d'un Do (2ème octave). 29
24 Représentation du chromagramme d'un extrait musical. 30
25 Vue schématique de la construction d'un vecteur chroma à partir d'un spectre d'amplitude. 31
26 Cycle des quintes présentant les classes de notes. 34
27 Les profils de tonalité de Krumhansl et de Temperley. 35
28 Modèle en spirale de Chew. 39
29 Espace basique de l'accord de Do Majeur dans la tonalité de Do Majeur. . . 40
30 Illustration de la distance entre $x = (C_M, C_M)$ et $y = (G_M, G_M)$. 41
31 Exemple de regroupement des notes suivant une représentation hiérarchique. 45
32 Exemple de représentation hiérarchique de la structure. 45
33 Exemple de réduction schenkerienne et de représentation hiérarchique de cette réduction. 47
34 Exemple de représentation d'élaboration à droite entre deux notes g et d, d'après la Théorie Génératrice de la Musique Tonale. 48
35 Analyse par Lerdahl et Jackendoff du début de la sonate pour piano en La de Mozart. 49
36 Modèle de l'harmonie de Bharucha. 50
37 Hiérarchie rythmique pour le groupement de notes. 51
38 Exemple de détection de tonalité d'après le modèle hiérarchique proposé par Rizo. 52
39 Arbre d'harmonie proposé, composé de cinq niveaux représentant cinq informations tonales différentes. 53
40 Exemple d'arbre d'harmonie représentant un extrait musical. 54
41 Exemple de représentation différente d'un arbre d'harmonie complet d'un extrait musical. 55
42 Exemple de transformation homorythmique. 58
43 Illustration du *spectral smoothness*. 61
44 Illustration de la construction d'un profil de notes. 65
45 Profils d'accords en triade de Do Majeur et Do mineur. 66
46 Comparaison de profils de notes d'une même mesure en Do Majeur selon leurs pondérations. 67
47 Résultats des calculs de corrélations. 68
48 Illustration d'un Modèle de Markov Caché (HMM). 73
49 Influence de l'ordre d'apparition des notes sur la tonalité perçue. 78
50 Profils d'intervalles de Madsen. 78
51 Profils de mode. 82
52 Les quatre premières mesures de *Little Wing* de Jimi Hendrix. 84

53 Profils de notes des premières mesures de *Little Wing* de Jimi Hendrix. . . . 84
54 Corrélations obtenues par comparaison entre profil de notes et profils de tonalités. . . . 86
55 Découpage de l'entrée musicale (symbolique ou audio) en segments et énumération de couples candidats. . . . 92
56 Illustration des couples candidats pour les différents segments de la pièce analysée. . . . 94
57 Illustration du graphe acyclique de couples candidats. . . . 94
58 Exemple de chemin optimal dans le graphe de couples candidats. . . . 95
59 Illustration du processus complet de sélection du chemin final dans le graphe de candidats énumérés. . . . 96
60 Estimation de similarité des distances musicales selon le rang moyen (en abscisse), comme proposé par le MIREX. . . . 102
61 Exemple d'extension de règle d'énumération causant une erreur dans l'estimation d'accords. . . . 103
62 Histogrammes représentant le nombre de morceaux par intervalle d'efficacité sur l'estimation des accords. . . . 106
63 Pondération des fréquences harmoniques pour le calcul des *HPCPs*. . . . 109
64 Processus d'obtention d'un chromagramme de *HPCPs*. . . . 109
65 Illustration de l'approche multi-résolutions. . . . 110
66 Illustration de l'utilisation des séquences d'accords pour la recherche de similarité dans une base de données. . . . 124
67 Illustration de la méthode *Tonal Pitch Step Distance*. . . . 125
68 Transcriptions manuelles d'extraits (correspondant aux motifs A et B) des deux morceaux *My Sweet Lord* (G. Harrison) et *He's So Fine* (R. Mack). . . . 136
69 Courts motifs musicaux composant la structure des deux morceaux *My Sweet Lord* (G. Harrison) et *He's So Fine* (R. Mack) : motif A (en haut), motif B (au centre) et motif C (en bas). . . . 137
70 Une représentation des accords (au dessus des paroles) utilisés pour la chanson *If I Ruled The World*, interprétée par Tony Bennett. . . . 142
71 Un extrait de la grille d'accords de *A Night In Tunisia*, de Dizzy Gillespie, appartenant au style be-bop. . . . 144
72 Une représentation des accords (au dessus des paroles) utilisés pour la chanson *The Thrill Is Gone*, de B.B. King, appartenant au style blues. . . . 144
73 Exemple d'utilisation des répétitions entre vecteurs chroma pour le système d'estimation d'accords. . . . 145

74 Différence entre les scores de l'estimation d'accords obtenus entre les configuration *optimum sans structure* et *décision* sur chacun des morceaux de la base de données. 147
75 Illustration des regroupements successifs de vecteurs chroma consécutifs en accords. 148

Remerciements

Je tiens ici à remercier l'ensemble de l'équipe Son du LaBRI pour son accueil chaleureux durant les trois années de cette thèse, ainsi que pendant les trois mois du stage de recherche qui les ont précédées. Un grand merci également à Pascal Ferraro, Julien Allali, Robert Strandh et Benjamin Martin avec qui j'ai eu l'occasion de travailler et d'échanger.
Je voudrais ensuite remercier tout particulièrement mes encadrants Myriam Desainte Catherine, Pierre Hanna et Matthias Robine, tout d'abord pour m'avoir proposé ce sujet de thèse et m'avoir fait part de leurs conseils avisés et de leur soutien tout au long de mon travail, mais également pour m'avoir réellement fait découvrir le monde de la recherche et de m'avoir donné l'envie de poursuivre dans ce domaine.
Enfin, je ne peux terminer ces remerciements sans citer l'aide précieuse apportée par Stéphanie, qui a su m'aider et m'accompagner pour terminer ces travaux.

Chapitre 1

Introduction

La musique occidentale se caractérise par son système de notation, qui la distingue de la musique à tradition orale. Pour pouvoir se transmettre, cette musique nécessite en effet une forme de représentation : un compositeur classique cherche par exemple, en composant une symphonie, à coucher sur papier une représentation de cette dernière, de manière à ce qu'un orchestre puisse ensuite rejouer sa création. La partition musicale est ainsi restée, durant plusieurs siècles, la forme principale de représentation musicale.
Depuis l'avènement des enregistrements numériques, datant de la seconde moitié du XXème siècle, la représentation la plus couramment utilisée pour un morceau de musique est son enregistrement en studio. Aujourd'hui, lorsque l'on écoute deux personnes comparer leurs avis sur un morceau de musique donné, c'est implicitement de l'enregistrement studio de ce morceau dont il est question.
Face à cette nouvelle forme de représentation musicale, il est notamment intéressant de chercher à en extraire une description, permettant par exemple aux apprentis musiciens de pouvoir réinterpréter certaines chansons de leurs artistes favoris. Une description particulièrement puissante de la musique pop actuelle consiste ainsi en des séquences d'accords. Un guitariste ayant écouté plusieurs fois *Heartbreaker*, de Pat Benatar, cherchera en effet à s'en procurer une description sous la forme d'une séquence d'accords, pour pouvoir reprendre ce titre, seul ou accompagné. La figure 1 illustre une telle représentation, très répandue sur Internet [1], et suscitant un vif intérêt de la part de la communauté des guitaristes.
Il est ainsi possible de décrire une chanson pop à l'aide de sa séquence d'accords. La question de l'automatisation du processus d'estimation des accords est donc posée, de même que l'automatisation de l'estimation de n'importe quelle autre description de la musique.

1. comme sur le site *911tabs* par exemple : `http://www.911tabs.com/`

```
Fm                      Ab                      Bb
You're a heartbreaker, dream maker, love taker,
                  C  C#    C
don't you mess around with me.
Fm                      Ab                      Bb
You're a heartbreaker, dream maker, love taker
                C  C#    C
don't you mess around with me.
```

FIGURE 1 – Une représentation des accords (au-dessus des paroles) pour la chanson *Heartbreaker* de Pat Benatar.

L'extraction d'informations musicales (ou MIR, *Music Information Retrieval*) [TSDD93] est un domaine interdisciplinaire, faisant intervenir la musicologie, la théorie de l'information, l'informatique et les sciences cognitives. Elle s'intéresse à la représentation, à la classification et à l'organisation de l'information et de la description musicale. Cette dernière peut mettre en jeu des paramètres très différents, tels que la temporalité, la hauteur, le timbre, ou encore l'harmonie. Ces paramètres n'étant pas indépendants les uns des autres, ils rendent d'autant plus complexes les problèmes posés par la recherche d'informations musicales. On trouvera sur la figure 2 une illustration du domaine de recherche MIR. On remarque la complexité des interactions entre les différentes facettes du MIR, laissant supposer la difficulté inhérente à l'étude de ce domaine. L'analyse des informations relatives à la théorie musicale (accords, tonalités, mélodie, tempo, etc) doivent en effet intégrer des éléments informatiques (recherche de motifs, approches statistiques), faire l'objet d'évaluation nécessitant des données annotées et être intégrée dans des méthodes de calcul de similarité.

L'extraction automatique d'informations musicales est un domaine de recherche présentant de grands enjeux, notamment à l'heure où le nombre de références musicales dans les catalogues en ligne connaît une réelle expansion. En 1999, on estimait à plus de 100000 le nombre de nouveaux titres rejoignant les catalogues musicaux chaque année [UZ99]. En 2011, une requête vide sur le catalogue en ligne d'Amazon [2] met en évidence plus de 17 millions de titres. Toutes ces pièces de musique font à l'heure actuelle l'objet de références sous forme d'étiquetages (*tags*), qui sont des descriptions textuelles du contenu des données pour en faciliter la classification. Parmi les étiquetages les plus répandus, on pourra citer :

- le nom de la pièce de musique (par exemple, "Let It Be"),
- l'interprète (par exemple, "The Beatles"),
- le(s) compositeur(s) (par exemple, "Lennon - McCartney"),
- la date de parution (par exemple, "May 1970").

2. le 17/11/2011, l'accès à la page `http://www.amazon.com/s?url=search-alias\%3Ddigital-music\&field-keywords=\&x=14\&y=8` montre un accès à 17 647 926 titres

Ce type de référence peut s'avérer utile, en particulier pour faire une recherche dans une grande base de données. Néanmoins, elle nécessite un étiquetage manuel, qui peut s'avérer d'autant plus long que la base de données comporte un nombre important de titres. Cet étiquetage peut également être subjectif, notamment conernant le style musical. De plus, un utilisateur souhaitant faire une recherche doit connaître une référence précise pour pouvoir trouver le fichier désiré. Enfin, un utilisateur souhaitant trouver un morceau similaire à un de ses morceaux favoris ne pourra tirer aucune utilité des étiquetages des morceaux, car la distance en terme d'étiquetage ne correspond pas à la différence musicale entre deux pièces de musique. Deux morceaux peuvent par conséquent présenter deux noms très proches sans pour autant être similaires musicalement parlant : ainsi, *We Will Rock You* de Queen ne présente aucune similarité musicale particulière avec *Rock With You* de Michael Jackson.

Face à de telles limites, il est souhaitable de décrire la musique non plus uniquement par des références textuelles (artiste, titre, ...), mais par des informations relatives au contenu de la pièce de musique concernée. En musique occidentale notamment, des paramètres tels que la mélodie, le rythme, l'harmonie, le timbre, ou encore la structure sont des informations relatives au contenu musical. De telles informations sont d'un niveau sémantique musical supérieur, et apparaissent donc pertinentes pour décrire le contenu musical d'une pièce de musique [Dow03].

1.1 Description des paramètres musicaux

La musique occidentale peut être principalement décrite par cinq paramètres musicaux [Duc03] : le rythme, le timbre, la forme (ou structure), la mélodie et l'harmonie. Il est à noter que certains de ces paramètres sont liés. Ainsi, la mélodie contient des informations relatives au rythme. De même, l'harmonie peut être influencé par la (ou les) mélodie(s).

1.1.1 Rythme

Le rythme est associé à l'idée de mesure temporelle dans la musique : d'après Hubbard, "toute succession de sons présente un rythme qui lui est propre et qui caractérise son mouvement temporel" [Hub08]. En musique occidentale, le rythme permet de déterminer une mesure de la durée d'une note par rapport aux autres. Le tempo d'un morceau peut être assimilé à une horloge interne, et associe une durée temporelle à une durée musicale de référence.

Le rythme est ainsi la seule composante musicale (avec la structure) dans laquelle la notion de hauteur des sons n'intervient pas. Il existe certaines pièces musicales (et même certains styles musicaux, comme la musique ewe[3]) dans lesquelles la seule composante

3. http ://en.wikipedia.org/wiki/Ewe_music

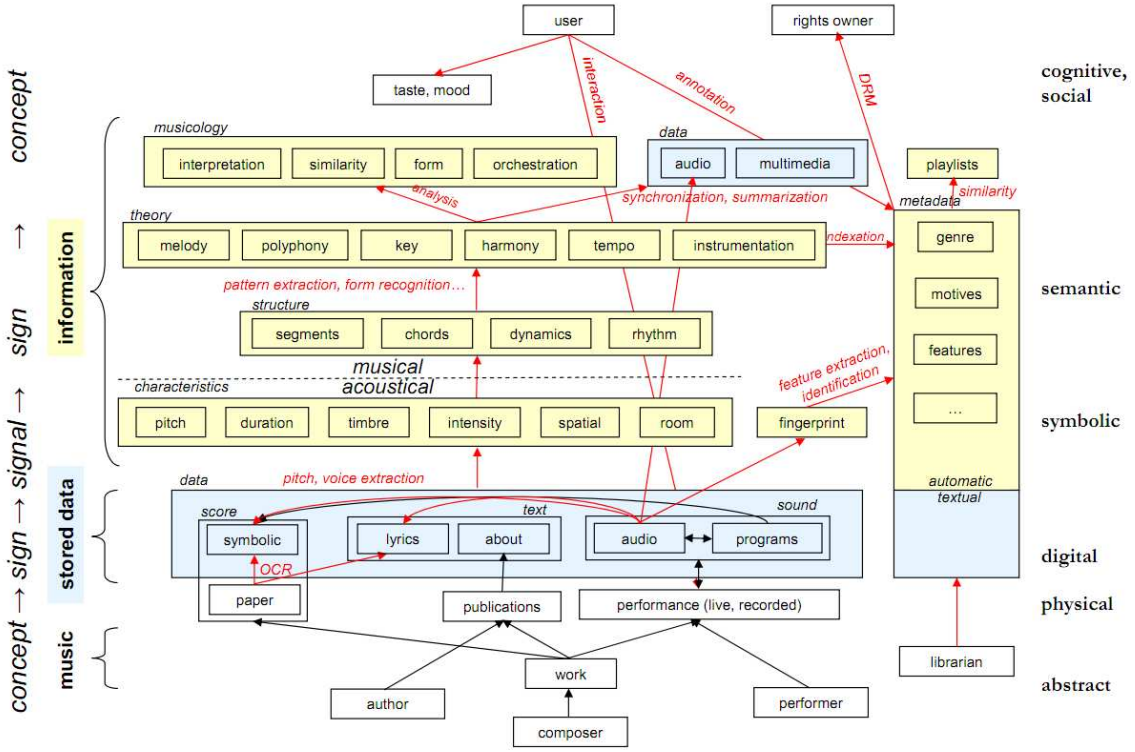


FIGURE 2 – Image représentative du domaine de recherche MIR (*Music Information Retrieval*), d'après [Fin04]

musicale utilisée est le rythme. Toutefois, dans la plupart des styles musicaux, et dans la musique occidentale en particulier, le rythme est combiné aux autres paramètres musicaux.

1.1.2 Timbre

Selon Jean-Claude Risset, "il existe plusieurs définitions du timbre : on le définit parfois comme l'attribut qui permet de différencier deux sons de même hauteur et de même durée ; ou encore comme l'attribut qui permet de reconnaître l'origine du son - et en particulier de distinguer des sons émis par des instruments de musique différents" [Ris78]. Le timbre d'un son musical peut être décrit par la répartition des pics de fréquence d'une fréquence fondamentale, ainsi que par leurs amplitudes respectives.

Plus simplement, le timbre est la composante musicale qui permet d'expliquer la différence ressentie lorsqu'une même phrase musicale est jouée successivement par deux instruments différents. Le timbre est ainsi une composante musicale singulière, qui ne peut pas être précisément transcrite sur une partition de musique traditionnelle.

1.1.3 Forme

La forme d'une pièce désigne l'agencement dans le temps des divers motifs musicaux la composant. Dans la littérature musicale, on peut trouver la forme musicale définie comme l'ensemble des stratégies permettant de trouver un juste milieu entre une répétition uniforme et une évolution uniforme [Sch38]. Dans la musique classique, la répétition uniforme est caractérisée par la forme strophique, qui consiste à répéter un certain nombre de fois la même cellule musicale (AAAA...). A l'inverse, l'évolution uniforme, appelée medley ou potpourri, compile différentes cellules musicales sans aucune répétition (ABCD...). Parmi les autres formes, toujours dans la musique classique, on pourra citer la forme binaire (AB), la forme ternaire (ABA), le rondo (symétrique : ABACABA ou asymétrique, ABACADAEA). Plus récemment, dans la musique moderne, on ne nomme plus les formes musicales dans leur ensemble, mais les motifs qui les composent. On parlera par exemple de thème, pont et chorus en jazz, ou de couplet, refrain et pont en musique pop.

1.1.4 Mélodie

Une mélodie désigne une succession de différentes hauteurs dans le temps. Cette succession est cohérente dans le sens où ces hauteurs sont perçues comme étant liées entre elles, et formant une même entité. [Ran03]. Il est également possible de définir une mélodie comme étant une entité musicale formée par une suite de notes jouées par un ensemble d'interprètes. Il arrive très fréquemment qu'une pièce comporte une mélodie répétée au cours de la pièce, formant ainsi un thème, qui peut être sujet à certaines variations.
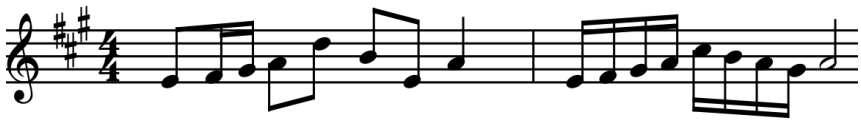


Figure 3 – Exemple de mélodie en La Majeur.

La mélodie peut être décrite comme l'aspect "horizontal" de la musique, en faisant se succéder les notes les unes aux autres dans le temps, alors que l'harmonie désigne la composante "verticale" de la musique, en faisant référence à l'effet produit par des notes jouées simultanément.

1.1.5 Harmonie

Au sens le plus général du terme, l'harmonie désigne une combinaison quelconque de plusieurs hauteurs [Hub08]. Plus précisément, elle fait référence aux différentes hauteurs entrant dans la composition d'un accord. L'harmonie constitue un aspect fondamental de la musique occidentale, et y est beaucoup plus développée que dans les autres musiques.

Au Moyen-Âge, le concept d'harmonie désignait l'emploi simultané de deux notes. Pendant la renaissance, les accords à trois sons devinrent la norme, et le restèrent jusqu'au XXème siècle. Des accords plus complexes firent alors leur apparition, répondant ainsi aux désirs de compositeurs comme Bartok ou Stravinsky de manipuler les tensions harmoniques issues d'accords jusque là peu employés. Cependant, les styles musicaux émergents au XXème siècle comme le blues, le jazz ou la musique pop emploient toujours des accords et une harmonie dans laquelle les accords mettant en jeu des empilements de tierces occupent une place prépondérante, que ce soit dans la composition ou dans l'arrangement.

L'harmonie ne peut pas être complètement dissociée des autres paramètres musicaux. L'harmonie se rapproche en effet du rythme, en créant des tensions tonales pour ensuite les résoudre, engendrant ainsi un mouvement rythmique vers l'avant. La durée des notes joue également un rôle car les notes jouées le plus longtemps correspondent souvent aux notes possédant un rôle particulier dans l'harmonie. L'harmonie entretient aussi une relation particulière avec la mélodie, dans le sens où les notes de la mélodie peuvent selon les cas affirmer ou à l'inverse contraster avec l'harmonie l'accompagnant, suivant ainsi les désirs du compositeur ou de l'arrangeur.

L'harmonie est donc un paramètre musical prépondérant dans la musique occidentale. Son étude permet d'appréhender les mécanismes de perception des tensions et des résolutions dont la musique est constituée. L'analyse de l'harmonie est ainsi une composante prépondérante du MIR, et permet une description particulièrement intéressante de la musique occidentale. C'est la raison pour laquelle nous avons ici choisi d'étudier les modèles et l'analyse de l'harmonie dans la musique occidentale

1.2 Résumés des chapitres

La présente étude est organisée comme suit :

Chapitre 1

Le chapitre 1 présente le contexte et les motivations des travaux, et notamment l'application de l'estimation des informations tonales à l'évaluation de la similarité musicale.

Chapitre 2

Ce chapitre propose une description des informations tonales étudiées, en distinguant notes, groupes de notes, accords, tonalités locales, et tonalité générale. Chacune de ces informations fait l'objet d'une définition et d'une modélisation informatique. Des notions de théorie musicale sont également introduites pour faciliter la compréhension du lecteur. Une présentation des descripteurs des informations tonales selon les représentations musicales (symbolique ou audio) est ensuite proposée.

Chapitre 3

Une fois les différentes informations tonales présentées, leurs différentes structurations au sein de modèles de l'harmonie sont étudiées. Les modèles de l'harmonie existants sont présentés, avant qu'une nouvelle proposition de modèle ne soit exposée. Ce nouveau modèle de l'harmonie est hiérarchique, et nommé arbre d'harmonie.

Chapitre 4

Suite à l'organisation des différentes informations tonales au sein d'un modèle de l'harmonie, nous nous penchons sur l'estimation de ces informations, en distinguant pour chacune les méthodes d'analyses depuis une représentation symbolique de la musique et depuis le signal audio. De nouvelles propositions d'amélioration pour les méthodes d'estimation des accords et des tonalités sont également exposées.

Chapitre 5

Ce chapitre présente une nouvelle méthode pour l'estimation conjointe des accords et des tonalités locales. Cette méthode est applicable depuis une représentation symbolique de la musique comme depuis le signal audio. Pour chacune de ces représentations de la musique, nous décrivons le fonctionnement de la méthode et proposons des évaluations de sa précision.

Chapitre 6

L'estimation de la similarité musicale, qui constitue une des principales applications de la modélisation et de l'estimation d'informations tonales, est étudiée dans ce chapitre. Les méthodes de calcul de similarité entre séquences d'accords sont d'abord présentées, avant que nous ne proposions une nouvelle méthode de calcul de similarité harmonique utilisant le modèle en arbre de l'harmonie.

Chapitre 7

Ce chapitre présente certaines pistes de recherche ouvertes par la présente étude. Pour finir, une conclusion propose un résumé des principales contributions de cette thèse.

Chapitre 2

Les informations tonales

Interrogé sur l'harmonie dans la musique occidentale, un musicien évoque "les notes qui s'enchaînent, formant des gammes qui affirment des modes, les accords formant des cadences, instaurant des tonalités articulées autour de modulations...". Ce vocabulaire, courant pour les musiciens, n'est pas précisément défini pour l'informaticien souhaitant modéliser ou estimer une note, un accord ou une tonalité. Ce chapitre présente différentes informations tonales (c'est-à-dire relatives à l'harmonie) et les moyens de les décrire, que ce soit depuis une représentation symbolique de la musique ou depuis le signal audio. Dans une première partie, nous présentons différentes informations tonales et mettons en avant la difficulté d'en obtenir des définitions claires et formelles. Nous proposons donc une paramétrisation de ces informations en vue d'établir, par la suite, un modèle de l'harmonie. Dans une seconde partie, nous présentons les deux descripteurs les plus utilisés pour ces informations tonales, qui seront employés ultérieurement dans les différentes méthodes d'analyse de l'harmonie proposées.

2.1 Définitions

La modélisation de l'harmonie d'une pièce de musique doit prendre en compte différents niveaux de description, pour décrire le plus précisément possible l'information tonale, qu'elle soit de *haut niveau* (tonalité) ou de *bas niveau* (notes de musique).

Pour ce faire, nous avons distingué et défini cinq informations tonales distinctes des notes de musique à la tonalité générale d'une œuvre. Il sera donc possible de choisir l'information offrant la meilleure description de l'aspect de l'harmonie auquel on s'intéresse, en fonction du but recherché. Ces différentes informations permettront également de faire un pas vers une représentation générale de l'harmonie.
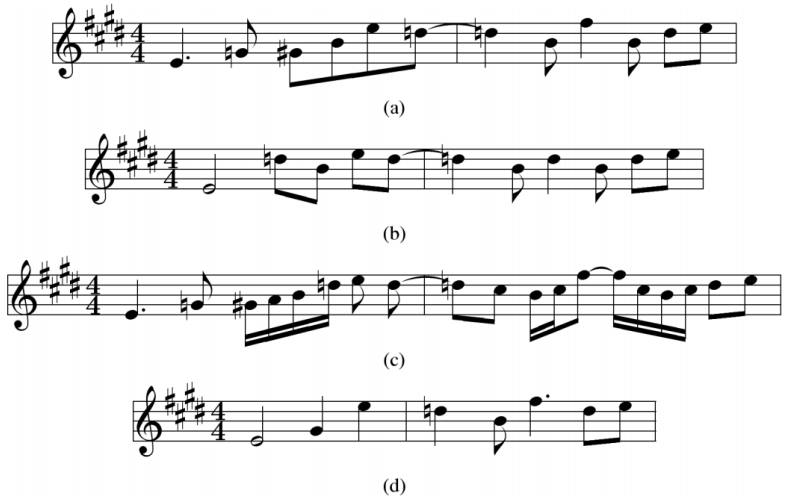


FIGURE 4 – Quatre variations sur *Day Tripper* des Beatles (d'après [RHF07]). L'harmonie reste la même (première mesure en Mi mineur, deuxième en Si mineur), alors que les notes diffèrent fortement entre les quatre extraits.

2.1.1 Note de musique

"Atome de la musique occidentale", d'après le musicologue Jean-Jacques Nattiez [Nat87], une note de musique est un symbole, utilisé pour former une partition. Elle est constituée d'une *hauteur* et d'une *durée*, et peut être annotée à l'aide d'informations telles que les nuances, les attaque, le vibrato... Les notes de musique constituent le niveau d'information le plus bas dans l'harmonie.

La seule étude des notes ne permet pas une analyse de l'harmonie très pertinente, car pouvant être considérée comme très bruitée : de nombreuses notes peuvent en effet être analysées comme des ornements ou des notes de passage, et peuvent donc jouer un rôle de *bruit* pour l'harmonie (voir figure 4).

2.1.1.1 Intervalle

Un intervalle désigne l'écart de hauteur entre deux notes. Mesuré dans la musique occidentale en demi-ton, un tel écart peut être :

- harmonique si les deux notes sont jouées simultanément,
- mélodique si les deux notes sont jouées l'une après l'autre.

La figure 5 référence les différents types d'intervalle.

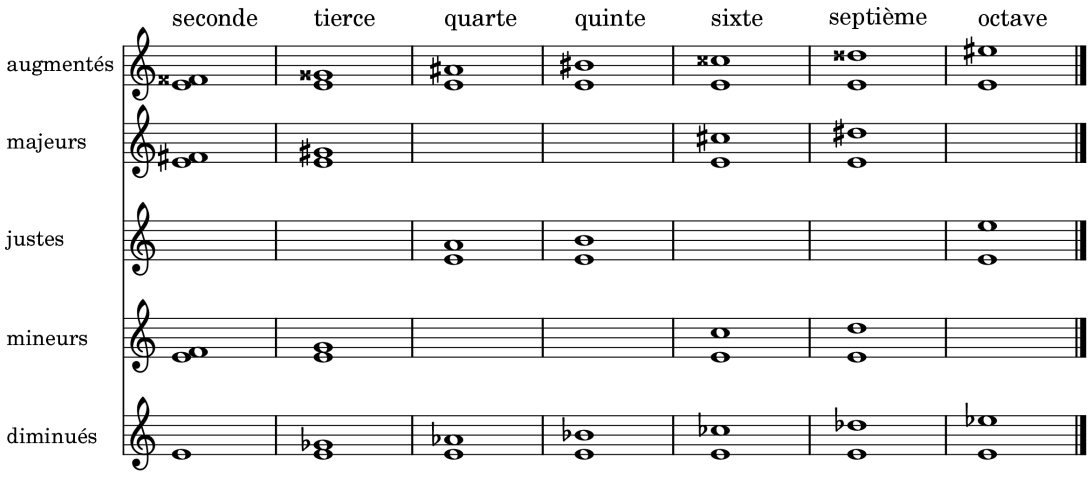


Figure 5 – Les intervalles musicaux, d'après [Wikd]

2.1.1.2 Note de passage et ornement

On qualifie de note de passage une note n'appartenant pas à l'harmonie locale (c'est-à-dire n'appartenant pas à la gamme courante) et située entre deux notes différentes de l'harmonie locale (voir figure 6). Un ornement est une note située au contraire entre deux notes semblables de l'harmonie locale.
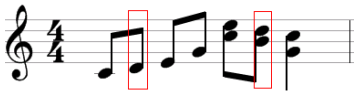


Figure 6 – Un exemple de notes de passage (notes encadrées).

2.1.1.3 Notation anglo-saxonne

Il est parfois fait référence aux notes de musique en utilisant la notation anglo-saxonne, en employant les lettres de l'alphabet. On trouvera dans le tableau 1 la correspondance entre le nom français et le nom anglo-saxon des notes de musique.

Nom français	do	ré	mi	fa	sol	la	si
Nom anglo-saxon	C	D	E	F	G	A	B

Table 1 – Noms français et noms anglo-saxon des notes de musique.

2.1.1.4 Paramétrisation

On définit les notes de musique à l'aide des paramètres suivants :

- *nom* : désigne la classe de la note, et est constitué du couple *hauteur* $\in$ {C, D, E, F, G, A, B}, et *altération* $\in$ {♮, ♭, ♯, ♭♭, ♯♯ ...},
- *octave* : marque la différence entre deux notes ayant le même nom, mais deux hauteurs distinctes (comme la première et la dernière note de la première mesure de l'exemple (d) de la figure 4),
- *date de début* : indique le temps auquel commence la note,
- *durée* : indique le temps pendant lequel la note est jouée.

Les paramètres *date de début* et *durée* peuvent être exprimés dans différentes unités, selon le type de représentation musicale étudié (secondes depuis le signal audio, et temps depuis une représentation symbolique, par exemple).
Il est à noter que d'autres paramètres, comme les nuances, peuvent également servir à caractériser une note de musique. Néanmoins, seuls les paramètres cités ci-dessus ont été utilisés dans la présente étude. Les nuances n'ont en effet qu'une influence mineure sur l'harmonie, comparée à l'influence des notes en elles-mêmes.

2.1.2 Groupe de notes

Un groupe de notes est constitué par un ensemble de notes sonnant en même temps. Deux notes peuvent toutefois être jouées durant une période commune de temps, sans pour autant débuter et finir toutes deux aux mêmes dates. Un problème est donc posé pour représenter un morceau comme une séquence de groupe de notes. Pour pouvoir prendre en compte les notes se chevauchant dans le temps, il est nécessaire de définir et d'appliquer une transformation illustrée par la figure 7. Cette transformation a été nommée *transformation homorythmique*.
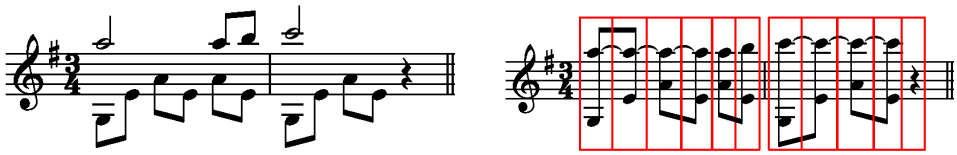


FIGURE 7 – Illustration de la transformation homorythmique. La figure de gauche représente un extrait polyphonique. La figure de droite représente le même extrait ayant subi une transformation homorythmique et faisant apparaître des groupements de notes (encadrés) servant à définir les groupes de notes.

Cette transformation préserve l'harmonie : les notes sont jouées de la même manière (il est impossible de distinguer à l'oreille une quelconque différence entre les deux extraits de la figure 7). De plus, cette transformation fait apparaître les notes sonnant en même temps sous forme de notes ayant une date de début et une durée communes.

2.1.2.1 Chiffrage

En harmonie classique, le chiffrage des groupes de notes désigne l'organisation des intervalles au-dessus de la basse, qui est la note la plus grave du groupe de notes.
On trouvera une illustration de chiffrage sur la figure 8 :

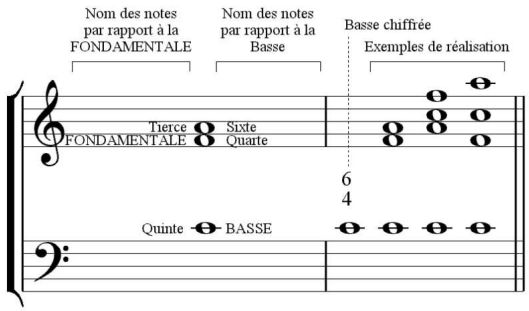


Figure 8 – Chiffrage de groupes de notes, d'après [Wikb].

Pour pouvoir chiffrer un groupe de notes, une seconde transformation est nécessaire. Cette dernière vise à supprimer les redondances formées par les notes apparaissant plusieurs fois au-dessus de la basse, et à rassembler toutes les notes du groupe de notes dans l'intervalle [*basse*, *basse à l'octave*]. On notera que cette seconde transformation conserve les propriétés tonales, toutes les notes jouées au moins une fois étant conservées (voir figure 9, qui présente en (c) la réduction de deux accords (a) et (b) possédant des notes apparaissant plusieurs fois).

2.1.2.2 Paramétrisation

On définit les groupes de notes à l'aide des paramètres suivants :
- *date de début* : indique le temps auquel débute les notes,
- *durée* : indique le temps pendant lequel les notes sont jouées,
- *basse* : désigne la nom de la note la plus grave de l'accord de notes (*hauteur*+*altération*, mais pas d'octave),
- *chiffrage* : désigne l'organisation des intervalles au-dessus de la *basse*,
- *fondamentale* : désigne le nom de la note sur laquelle est fondée la superposition de tierces de l'accord de notes (*hauteur*+*altération*, mais pas d'octave).

On note que le chiffrage, la basse et la fondamentale permettent d'avoir une représentation unique d'un groupe de notes d'un point de vue harmonique. En effet, deux groupes de notes présentant la même *fondamentale*, la même *basse* et le même *chiffrage* seront harmoniquement équivalents entre eux (voir figure 9).

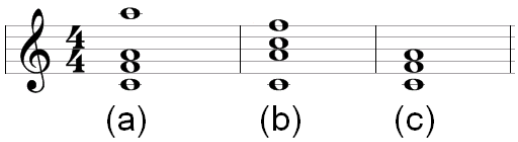


FIGURE 9 – Équivalence entre groupes de notes. Bien que composés de notes différentes, les groupes de notes (a) et (b) sont tous les deux représentés sous la forme (c), avec une basse de Do, une fondamentale de Fa, et un chiffrage 6-4.

2.1.3 Accord

Selon [Ran03], un accord désigne "un ensemble de trois notes (ou plus) sonnant simultanément, ou fonctionnant comme si elles sonnaient simultanément". Cette définition permet beaucoup d'interprétations différentes. La première partie de la définition correspond à ce que nous avons désigné comme étant un groupe de notes, en section 2.1.2. Nous allons donc nous appuyer sur la seconde partie de la définition ci-dessus pour caractériser ce qu'est un accord, et nous pencher sur ces notes qui "fonctionnent comme si elles sonnaient simultanément".
L'accord constitue une abstraction par rapport au groupe de notes. Les notes d'accords ne correspondent en effet pas nécessairement aux notes jouées par les interprètes d'un morceau, mais aux notes perçues comme prépondérantes par l'auditeur pendant une certaine période de temps. Cette durée peut varier, selon les pièces et les styles de musique, entre une fraction de seconde et une dizaine de seconde, et correspond à la durée de l'accord. Si on se réfère à la manière dont sont généralement représentés les accords, on remarque que s'il est possible de déterminer quelles notes les composent, aucune indication n'est en revanche mentionnée quant à la manière de les jouer. Ainsi, sur la figure 10, on remarque que le premier accord à jouer est un Si mineur, mais rien n'est indiqué quant à l'octave, le rythme ou le nombre des notes devant être jouées. Un accord marque ainsi un contexte tonal local, et non un ensemble de notes réellement notées et jouées.

À l'état fondamental, un accord est représenté sous la forme d'une superposition de tierces (voir figure 11). Un accord possède un degré, généralement noté en chiffre romain, qui qualifie l'intervalle entre la fondamentale de l'accord et la tonique de la tonalité. Ainsi, pour une pièce en Do Majeur, un accord de Do Majeur aura pour degré I alors qu'un accord de Sol Majeur aura pour degré V.

```
Bm                          F#
On a dark desert highway, cool wind in my hair
A                      E
Warm smell of colitas rising up through the air
G                           D
Up ahead in the distance, I saw a shimmering light
Em
My head grew heavy and my sight grew dim
F#
I had to stop for the night
```

Figure 10 – Une représentation courante des accords (au dessus des paroles) pour la chanson *Hotel California* du groupe The Eagles. Ce type de notation est très couramment rencontré dans le milieu de la musique pop.

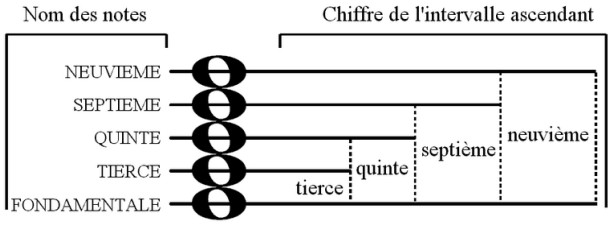


Figure 11 – Le nom des notes dans un accord de neuvième, d'après [Wika].

Sur les partitions, l'accord est fréquemment indiqué par le nom de sa fondamentale, et sous la forme d'une codification simplifiée.

Exemples : `DO` = do, mi, sol ; `DOm` = do, mi♭, sol ; `DOm7` = do, mi♭, sol, si♭ ; `DO7` = do, mi, sol, si♭ ; `DO9` = do, mi, sol, si♭, ré ; etc. Dans de nombreux styles musicaux (jazz, blues, rock, pop,...), les lettres anglo-saxonnes se substituent aux noms des notes latines. Ainsi Do Majeur s'écrit `C` ; Fa septième de dominante, `F7` ; Ré mineur, `Dm` ; Mi Majeur septième majeure, `EMaj7`, etc

2.1.3.1 Paramétrisation

On définit un accord à l'aide des paramètres suivant :

- *date de début* : indique le temps auquel début l'accord,
- *durée* : indique le temps pendant lequel l'accord est joué,
- *fondamentale* : désigne le nom de la note sur laquelle est fondée la superposition de tierces des notes composant l'accord,
- *type* : représente l'ensemble des intervalles présents dans l'accord. La liste des différents types d'accords peut s'avérer extrêmement longue, comme on peut le constater sur la figure 12. D'un point de vue combinatoire, il existe 220 types d'accords différents composés de 3 notes, 495 de 4 notes et 792 de 5 notes.

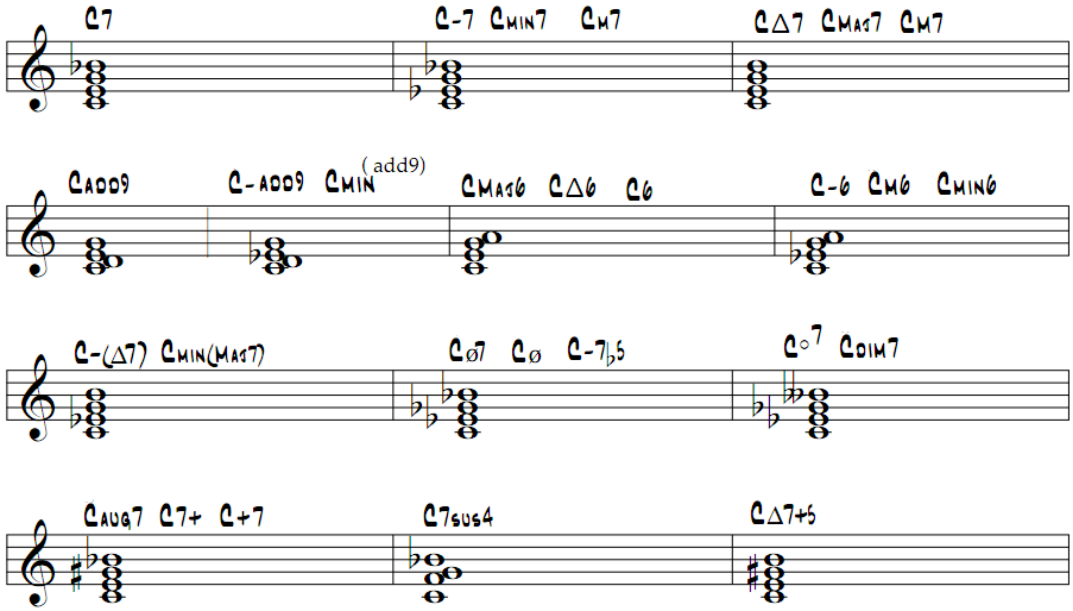


FIGURE 12 – Un extrait des différents accords de Do existants, d'après [Rea05].

2.1.3.2 Différence entre *groupe de notes* et *accord*

Tel qu'il a été défini, un *groupe de notes* constitue un ensemble de notes jouées simultanément, chaque note de ce groupe ayant la même durée. Ces notes sont jouées par les interprètes de la pièce, et apparaissent sur la partition.
Un *accord*, tel qu'il a été défini, doit davantage être vu comme un élément d'une progression harmonique, qui marque le contexte tonal prédominant sur une durée définie (souvent de l'ordre de la mesure ou demi-mesure).
La figure 13 illustre la différence entre un groupe de notes et un accord.

2.1.4 Tonalité locale

La tonalité locale de l'extrait d'un morceau peut être définie comme la tonalité principale de ce morceau s'il était réduit à l'extrait considéré. La séquence des tonalités locales forme les modulations d'une pièce de musique. Un exemple de modulation célèbre dans la musique pop est donné par le passage du couplet au refrain dans *Lucy In The Sky With Diamonds* des Beatles, avec le passage du couplet (en La Majeur) au refrain (en Sol Majeur).

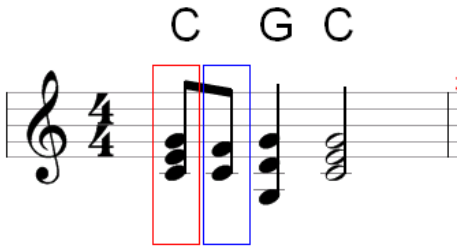


Figure 13 – Exemple illustrant la différence entre un groupe de notes et un accord. Sur cet exemple, on dénombre trois accords correspondant aux annotations C (Do Majeur), G (Sol Majeur) puis de nouveau C (Do Majeur) mentionnées au-dessus de la portée. Les groupes de notes constituant le premier accord de Do sont encadrés en rouge et en bleu, et sont (do mi sol) et (do fa). Or fa n'appartient pas à l'accord de Do Majeur (c'est une note de passage), mais le groupe de notes dont il est issu appartient à l'accord de Do majeur.

Il est possible d'évaluer la richesse harmonique de la pièce étudiée en la rapprochant du nombre de modulations utilisées. Néanmoins, une telle étude s'avérerait sans doute plus pertinente dans le cadre de pièces de musique classiques que dans le cadre de musiques de variétés. Ces dernières comprennent en effet un nombre de modulations généralement bien inférieur au nombre de modulations dans des pièces classiques (une multitude de chansons *pop* ne comportent d'ailleurs aucune modulation).

2.1.4.1 Paramétrisation

On définit une tonalité locale comme présentant les paramètres suivants :
- *date de début* : indique le temps auquel la tonalité débute,
- *durée* : indique le temps pendant lequel la tonalité s'étend,
- *tonique* : formée d'un *nom* (*hauteur*+*altération*) mais pas d'une *octave*, elle désigne le premier degré de la gamme de la tonalité,
- *mode* : majeur ou mineur.

2.1.5 Tonalité principale

Selon David Temperley [Tem99a], la tonalité fournit "le cadre dans lequel les notes et les accords sont perçus". Le concept de tonalité, comme celui d'accord, ne supporte pas de définition formelle et unique. Il renvoie à l'idée de stabilité harmonique, et plus particulièrement au fait qu'une note (la *tonique*) reste *stable* (ou *importante*) durant une certaine période de temps, qui est communément plus longue que celle de la durée d'un accord [Vos00]. Le choix de la tonalité d'un morceau de musique était très important durant la période classique : Steblin [Ste96] a ainsi montré que le choix de Do Majeur

symbolisait pureté, innocence et simplicité, alors qu'un choix comme La♭ mineur reflétait plutôt l'éternité, le jugement ou la mort.

2.1.5.1 Gamme et mode

Une gamme est un ensemble de notes qui sert de support à une tonalité. Elle présente une succession de notes, la première d'entre elle étant appelée la tonique. La distribution des intervalles formés par ces notes constitue le mode. Il existe un grand nombre de modes distincts. On citera notamment les modes majeurs, mineurs harmoniques, mineurs mélodiques, chromatiques, par tons et pentatoniques. Les trois modes les plus fréquents en musique occidentale sont illustrés sur la figure 14. On notera que le terme "mineur" désigne par défaut le mode mineur harmonique.

2.1.5.2 Cycle des quintes

Les différentes tonalités peuvent être organisées autour d'un cercle appelé cycle des quintes (représenté sur la figure 15).

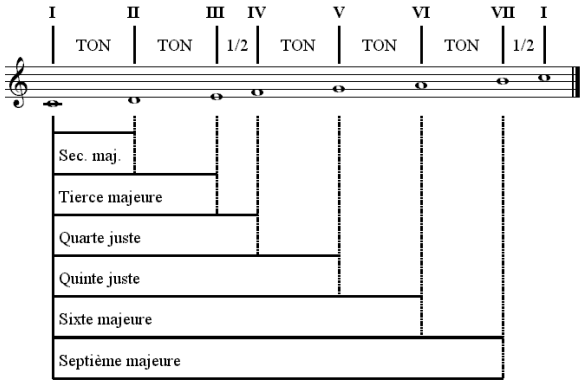
Cette représentation est notamment utile pour illustrer les concepts de ton relatif et de ton voisin. Deux tonalités sont relatives si et seulement si l'une est majeure, l'autre est mineure, et la tonique de la tonalité mineure se situe une tierce mineure en dessous de la tonique de la tonalité majeure (exemple : Do Majeur et La mineur). Deux tons relatifs partagent en outre la même armure, c'est-à-dire le nombre d'altérations présentes naturellement dans la gamme associée à la tonalité. Deux tonalités sont voisines si et seulement si elles ont le même mode (majeur ou mineur) et l'intervalle entre les toniques des deux tonalités est une quarte juste ou une quinte juste. Chaque tonalité possède ainsi deux voisines (par exemple, Do Majeur est voisin avec Fa Majeur et Sol Majeur).

2.1.5.3 Paramétrisation

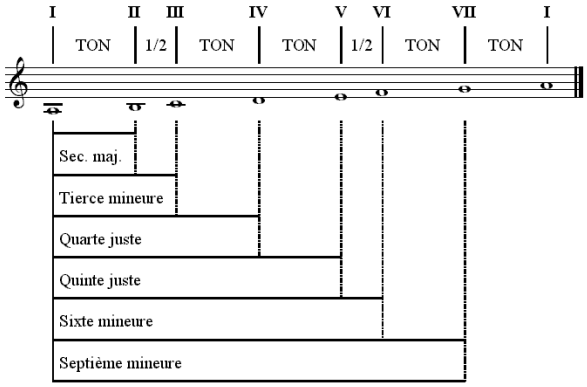
Une tonalité possède les mêmes paramètres qu'une tonalité locale, à l'exception de la date de début et de la durée, car cette tonalité ne varie pas au cours du morceau :

- *tonique* : formée d'un *nom* (*hauteur*+*altération*) mais pas d'une *octave*, elle désigne le degré fondamental (ou la note principale) de la gamme induite par la tonalité (voir section 2.1.5.1),
- *mode* : majeur ou mineur.

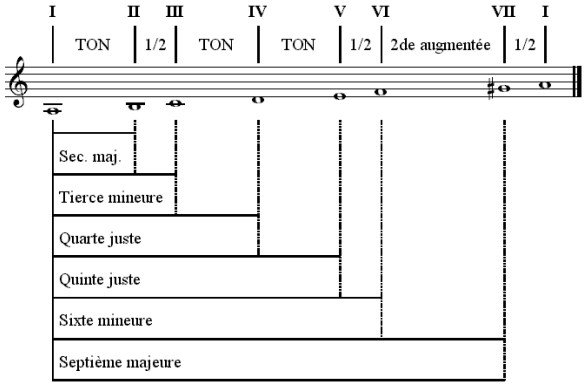
On trouvera sur la figure 16 un récapitulatif des informations tonales décrites dans ce chapitre, ainsi que leurs paramètres et une possible représentation pour chacune d'entre elles.



Le mode majeur



Le mode mineur naturel



Le mode mineur harmonique

Figure 14 – Les trois gammes de Do correspondant aux trois modes les plus couramment rencontrés.

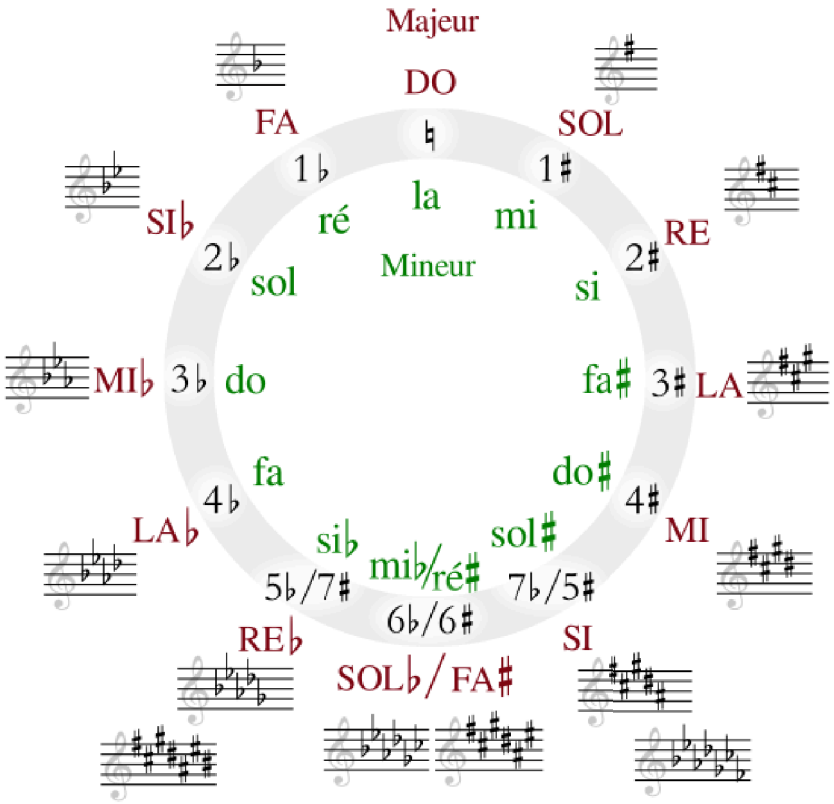


FIGURE 15 – Les différentes tonalités organisées autour du cycle des quintes, d'après [Wikc].

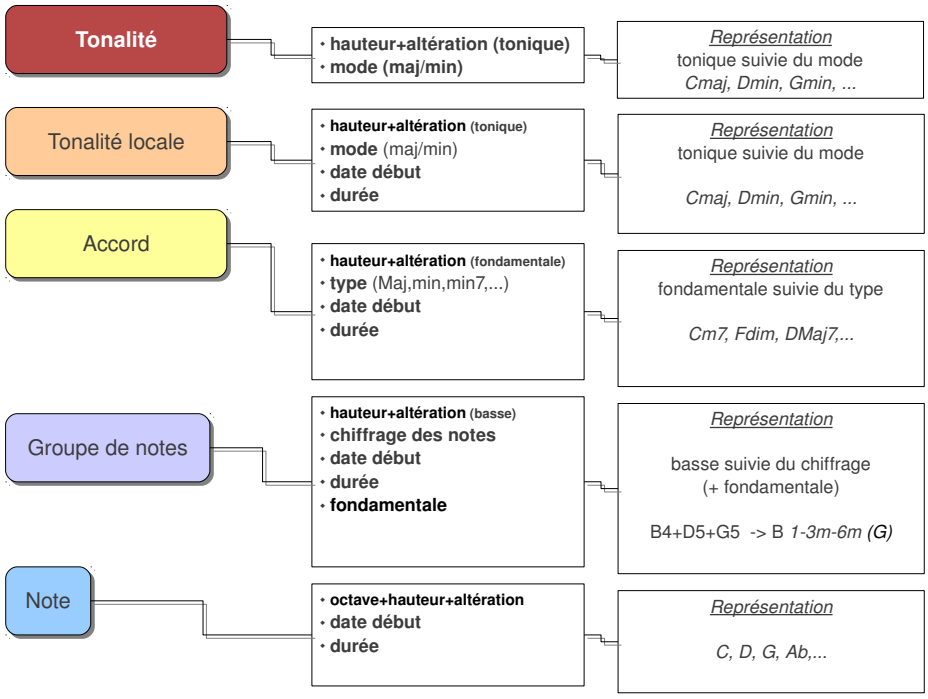


Figure 16 – Récapitulatif des informations tonales et de leurs paramètres.

2.2 Descripteurs des informations tonales

Après avoir décrit les différentes informations tonales, nous nous penchons sur la manière de décrire l'harmonie, pour deux types de représentation musicale : la représentation symbolique de la musique et le signal audio.

2.2.1 Depuis une représentation symbolique de la musique

2.2.1.1 Description

La représentation symbolique de la musique désigne toute forme de représentation où apparaissent les notes jouées par les différents instruments (ou interprètes) de la pièce de musique. La représentation symbolique la plus répandue et la plus populaire est la partition de musique. En informatique, le format MIDI (*Musical Instrument Digital Interface*) constitue également une représentation symbolique de la musique.

Dans un fichier MIDI, les différents instruments sont représentés par différentes pistes, chaque piste comportant une liste d'événements (voir figure 17). Bien que certains de ces événements peuvent coder des changements d'attaque, des nuances ou encore des effets, la plupart représentent des notes. En codant la hauteur de ces notes, leurs dates de début et de fin, on peut ainsi se ramener à une représentation appelée *pianoroll* (voir figure 18).

```
Note On, Channel:1, Key: 60 (C4), Velocity: 69
Note On, Channel:1, Key: 67 (G4), Velocity: 67
Note On, Channel:1, Key: 69 (A4), Velocity: 77
Note Off, Channel:1, Key: 60 (C4), Velocity: 71
Note On, Channel:1, Key: 60 (F4), Velocity: 65
Note Off, Channel:1, Key: 67 (G4), Velocity: 56
Controller, Channel:1, Number 74 (Sound Brightness), Value: 55
Controller, Channel:1, Number 71 (Sound Timbre), Value: 28
Pitch Wheel, Channel:1, Value: 3048
```

FIGURE 17 – Différents événements MIDI obtenus lors de la description textuelle d'un fichier MIDI.

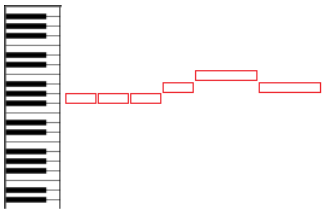


FIGURE 18 – Représentation en *Piano Roll*, correspondant au début de la mélodie populaire *Au Clair De La Lune*. Le temps apparaît de gauche à droite sur la figure.

2.2.1.2 Profils de notes

Depuis une représentation symbolique de la musique, il est possible d'utiliser un descripteur de l'harmonie qui prend en compte les différentes notes jouées dans la pièce de musique étudiée, et qui permet en particulier d'observer la fréquence d'apparition des différentes notes qui la composent. En effet, une pièce en Do Majeur, par exemple, aura tendance à privilégier l'utilisation des notes de la gamme de Do Majeur, et en particulier la tonique (Do), la dominante (Sol), et la tierce majeur (Mi), qui affirme le mode.
Le descripteur de l'harmonie le plus utilisé, et qui présente ces caractéristiques, est le profil de notes. Introduit pour la première fois par Krumhansl [KK82] dans le cadre d'expériences psycho-acoustiques, le profil de notes a ensuite été repris dans plusieurs études sur l'analyse de l'harmonie depuis une représentation symbolique de la musique [Tem99a, PB02]. Un profil de notes mesure l'influence des notes jouées par les instruments (ou interprètes) sur une période de temps donnée. Ces notes sont regroupées en 12 classes (voir tableau 2).

C	C#/Db	D	D#/Eb	E	F	F#/Gb	G	G#/Ab	A	A#/Bb	B

TABLE 2 – Les 12 classes de notes représentant les 12 dimensions d'un profil de notes.

Un profil de notes est donc un vecteur de dimension 12. Sa construction consiste à considérer chacune des notes jouées pendant l'intervalle de temps étudié, et à incrémenter la composante du vecteur correspondant à la classe de la note rencontrée. On peut soit incrémenter cette composante d'une unité pour chaque note correspondante, ou alors l'incrémenter d'une valeur relative à la durée de la note considérée, pour offrir une description plus précise de l'influence de cette note sur l'harmonie. Dans ce dernier cas, nous parlerons d'un profil de notes pondéré. Dans la suite de cette étude, tous les profils de notes seront supposés pondérés.

La figure 19 représente le début du menuet BWV-114 de J.S. Bach. La tonalité de ce menuet est Sol Majeur.
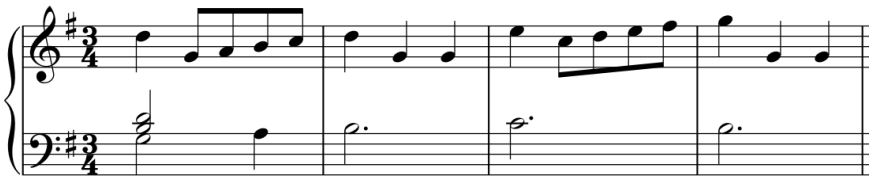


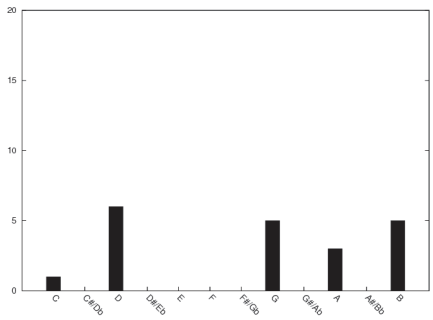
Figure 19 – Premières mesures du menuet BWV-114 de J.S. Bach.

La figure 20 représente les 4 profils de notes correspondant respectivement aux 4 premières mesures de l'extrait, ainsi que le profil de notes correspondant à l'extrait entier (qui s'obtient en sommant les 4 autres profils). Cette figure nous permet de constater que sur la totalité de l'extrait, les trois notes les plus influentes sont le Si (B), le Sol (G) et le Ré (D), qui correspondent respectivement à la médiante, la tonique et à la dominante de Sol Majeur, qui est précisément la tonalité de l'extrait. On constate également que cette tendance est conservée sur trois des quatre profils des quatre premières mesures. La seule exception est la mesure 3, pour laquelle le Do (C) et le Mi (E) sont prépondérants. Or, cette mesure est la seule des quatre mesures à ne pas comporter un accord de Sol Majeur, mais bien un accord de Do Majeur, dont Do et Mi sont respectivement la tonique et la médiante. Ce simple exemple nous permet donc de constater que le profil de notes est un descripteur de l'harmonie très intéressant, pourvu que l'on maîtrise la période de temps sur laquelle l'observation est réalisée.

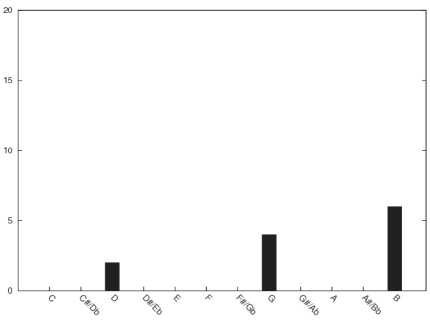
2.2.2 Depuis le signal audio

2.2.2.1 Description

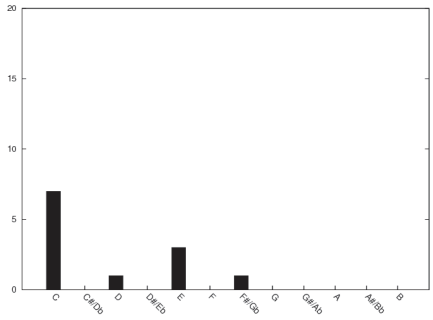
Nous nous intéressons ici à la représentation informatique du signal audio. Un tel signal est numérique, résultant de la quantification et de l'échantillonnage de l'onde sonore analogique. On trouvera une illustration de ces transformations sur la figure 21. L'échantillonnage consiste à capturer l'amplitude de l'onde sonore à intervalles réguliers, appelés



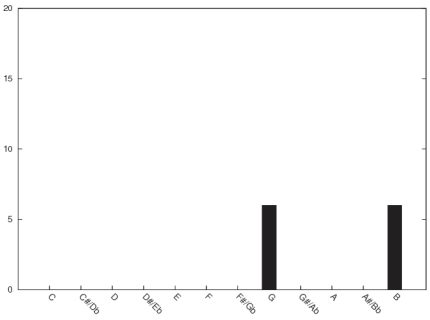
profil de notes pour la mesure 1



profil de notes pour la mesure 2



profil de notes pour la mesure 3



profil de notes pour la mesure 4

profil de notes pour la mesure 5

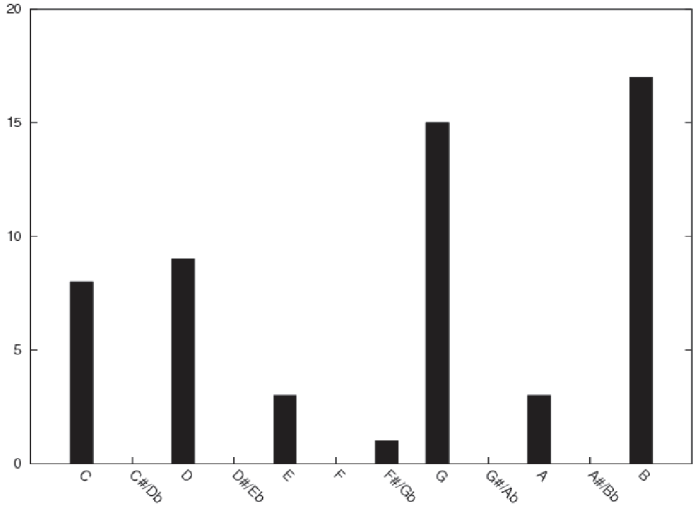


FIGURE 20 – Les profils de notes respectifs des 4 premières mesures du menuet BWV-114de J.S. Bach, suivis du profil de notes sur l'extrait entier. En ordonnée figure la pondération (durée cumulée des occurrences des classes de notes, en abscisse). La croche a été choisie comme unité de durée.

périodes d'échantillonnages (T_{ech} sur la figure 21). Plus ces périodes de temps sont rapprochées, plus la fréquence (inverse de la période) est importante, et meilleur sera l'échantillonnage (car plus fidèle). Le signal est donc représenté par une série de valeurs, appelés échantillons. La quantification consiste à encoder la valeur des échantillons à l'aide d'un nombre de bits donnés. Plus ce nombre de bits est grand, plus le pas de résolution (Q sur la figure 21) est petit, permettant ainsi une meilleure résolution pour les valeurs d'amplitude du signal numérisé.

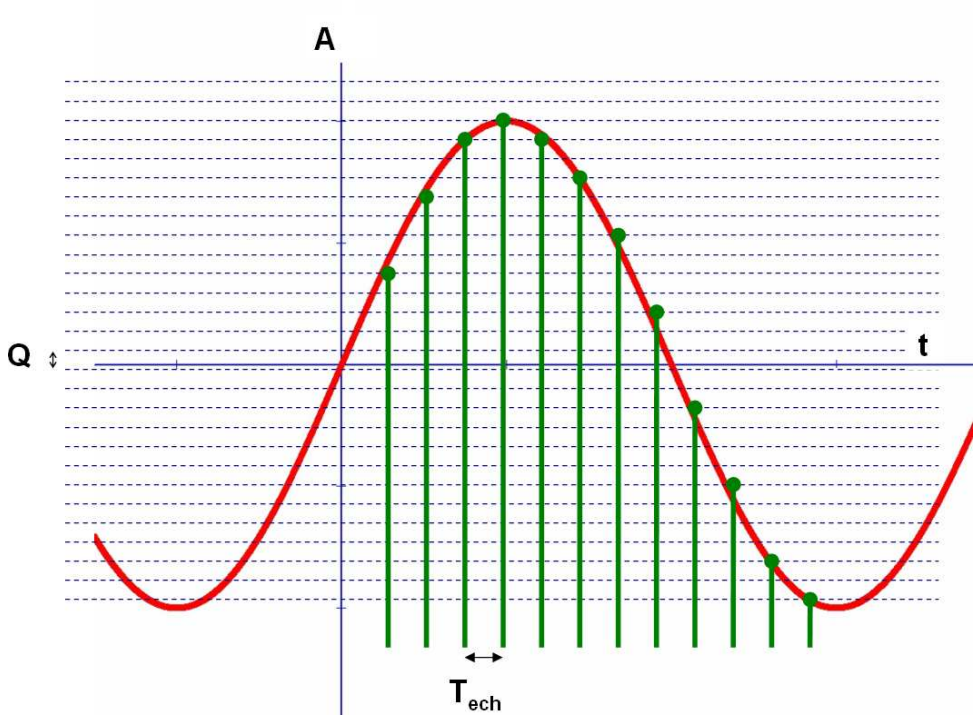


Figure 21 – Illustration de l'échantillonnage et de la quantification de l'onde sonore (en rouge), dans un processus de numérisation. L'amplitude du signal est en ordonnée, le temps en abscisse. T_{ech} désigne la période d'échantillonnage, soit la période de temps entre deux captures consécutives (en vert) de l'amplitude sonore.

Pour parvenir à une description musicale de l'harmonie d'un signal audio, il est nécessaire de faire intervenir les hauteurs jouées dans l'extrait étudié. La notion de hauteur est un paramètre perceptif lié au paramètre de fréquence. Une note de musique, lorsqu'elle est jouée par un instrument de musique, est émise sous la forme d'un ensemble de fréquences f_i, parmi lesquelles se trouve une fréquence particulière appelée fréquence fondamentale, et notée f_0. Cette fréquence est généralement la plus basse des fréquences f_i, et possède, la plupart du temps, l'amplitude la plus importante. La hauteur perçue par l'auditeur dépend de la valeur de f_0.

À l'aide de la transformée de Fourier, il est possible de connaître les fréquences des sons constituant un extrait musical. La transformée de Fourier d'une fonction continue f s'exprime mathématiquement comme la fonction F telle que :

$$F(x) = \int_{\mathbb{R}} f(t)e^{-i2\pi xt}dt$$

Pour l'appliquer à un signal s réel discret, en considérant N échantillons, on considère la transformée de Fourier discrète S, définie par :

$$S(k) = \sum_{n=0}^{N-1} s(n)e^{-i2\pi k\frac{n}{N}}, 0 \leq k < N$$

La figure 22 présente un signal audio numérique en représentation temporelle, et en dessous, une représentation spectrale où les différentes fréquences composant l'extrait sonore sont représentées en fonction du temps.

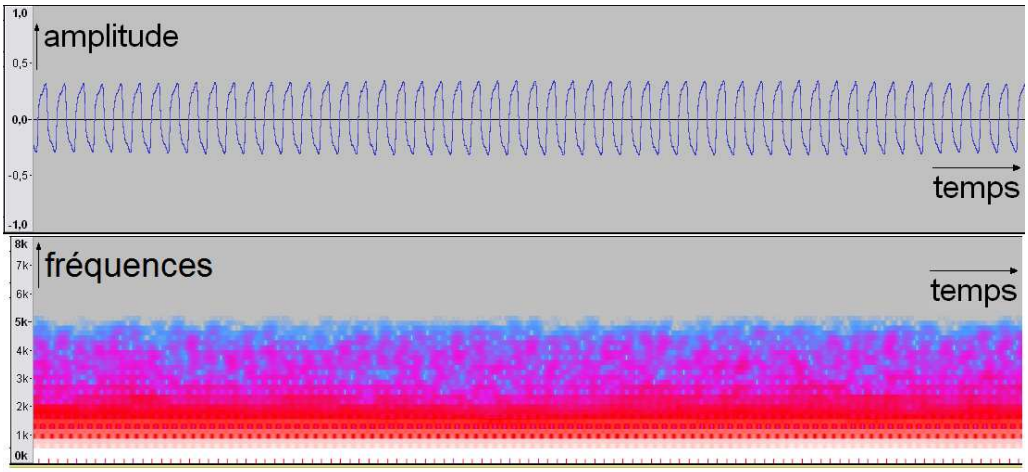


FIGURE 22 – Un signal audio représenté en amplitude en fonction du temps (en haut) et en fréquences en fonction du temps (en bas).

Cette décomposition en fréquences, qui intervient sur une *trame* de signal (un extrait de signal dont le début et la durée sont à définir), permet d'obtenir des informations sur les notes jouées pendant la trame considérée. En effet, chaque note de musique possède une fréquence fondamentale qui lui est propre. On trouvera dans le tableau 3 une correspondance entre les 12 demi-tons de la gamme chromatique sur trois octaves et leurs fréquences fondamentales respectives.

Comme il a été mentionné précédemment, lorsqu'une note n est jouée par un instrument réel, la fréquence f_0 n'est pas la seule à être émise. D'autres fréquences, f_i sont ainsi émises, et correspondent aux fréquences fondamentales d'autres notes de musique n_i. Les

octave	Fréquence fondamentale(en Hertz)											
	C	C#/Db	D	D#/Eb	E	F	F#/Gb	G	G#/Ab	A	A#/Bb	B
octave 3	130.8	138.5	146.8	155.5	164.8	174.6	184.9	195.9	207.6	220.0	233.1	246.9
octave 4	261.6	277.12	293.7	311.1	329.6	349.2	370.0	392.0	415.3	440.0	466.2	493.9
octave 5	523.2	554.3	587.3	622.2	659.2	698.4	739.9	783.9	830.6	880.0	932.3	987.7

Table 3 – Correspondance entre différentes notes et leurs fréquences fondamentales associées.

fréquences f_i et les notes n_i sont appelées *harmoniques* de la note n. La figure 23 représente les 10 premières harmoniques d'un Do grave (encadré en rouge sur la figure). La présence de ces harmoniques explique, d'un point de vue fréquentiel, la difficulté de l'analyse de la musique depuis le signal audio. Lorsque les fréquences $f_1 = 130$Hz et $f_2 = 392$Hz sont détectées, par exemple, il est difficile de déterminer si ces fréquences sont deux fréquences fondamentales, et que les notes Do et Sol sont jouées, ou si f_2 est une harmonique de f_1, et donc que seule la note Do est jouée.
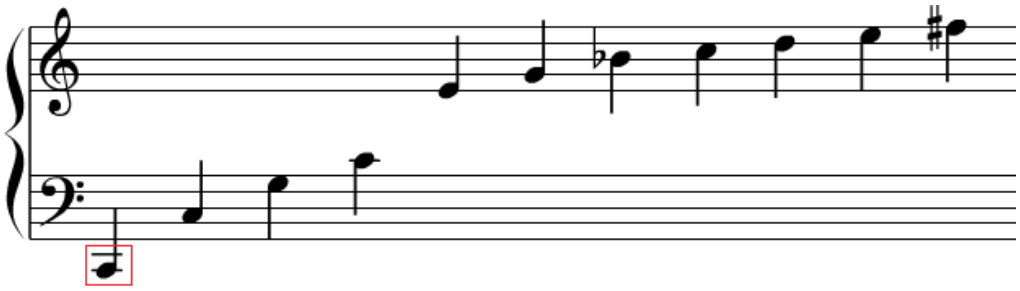


Figure 23 – Les 10 premières harmoniques d'un Do (2ème octave).

2.2.2.2 Vecteurs chroma

Dans le but d'utiliser un descripteur semblable aux profils de notes, mais utilisable depuis le signal audio, les vecteurs chromas (ou PCPs, *Pitch Class Profile*) ont été introduits en 1999 par Fujishima et Wakefiled [Fuj99, Wak99]. Leur construction utilise les informations spectrales associées à une trame de signal donnée pour construire un vecteur à 12 dimensions mesurant l'influence des 12 classes de notes, de la même manière qu'un profil de notes. On trouvera sur la figure 24 une illustration d'un chromagramme, qui présente différents vecteurs chromas au cours du temps.

Fujishima [Fuj99], puis Goto [Got06] et Ellis [EP07] proposent ainsi une construction des chromas illustrée sur la figure 25, consistant à rassembler les parties du spectre contenant des informations communes à la même classe de note. Gòmez, en 2006 [Góm06], propose une prise en compte des harmoniques des notes jouées et introduit les *HPCPs* (*Harmonic Pitch Class Profiles*).

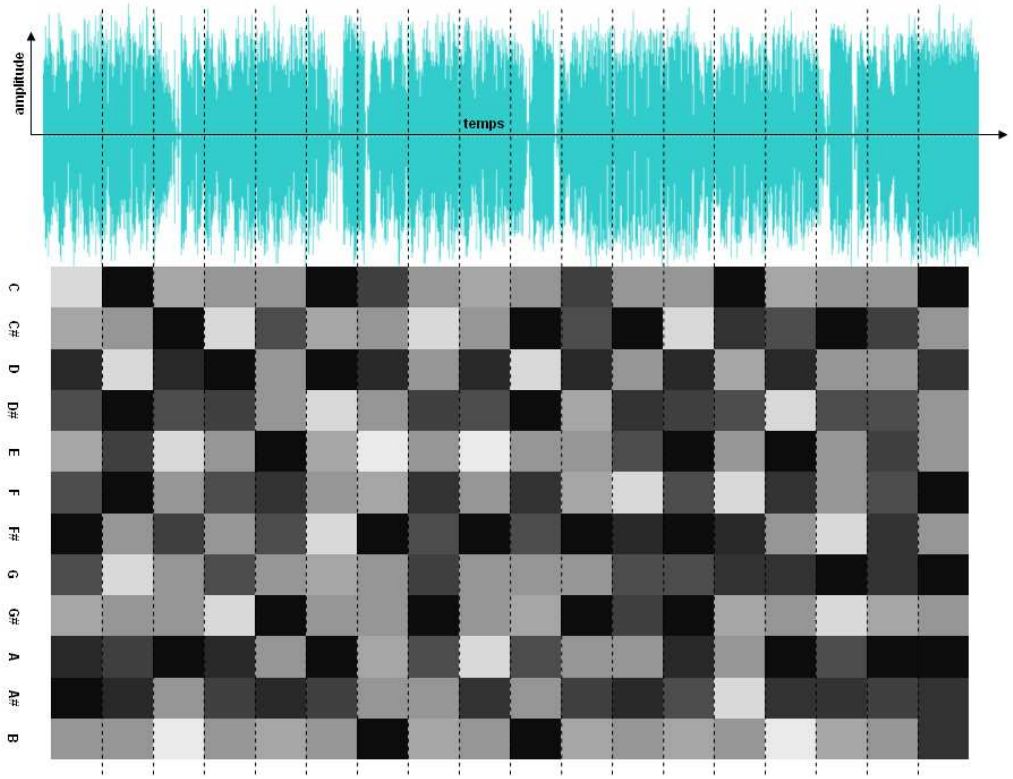


FIGURE 24 – Représentation du chromagramme d'un extrait musical.

D'autres méthodes peuvent être utilisées pour obtenir des vecteurs chromas, en utilisant par exemple une transformée Q constante [BP05, HS05], une estimation de fondamentale polyphonique [RK08, VPM08] ou encore un banc de filtres [MKC05].

Les vecteurs chromas sont ainsi comparables aux profils de notes car ces deux descripteurs mesurent l'influence des 12 différentes classes de notes sur un extrait musical. Là encore, la période d'observation (ou fenêtre temporelle) servant de support à l'estimation d'un vecteur chroma devra être ajustée selon l'information tonale recherchée. Les chromas se distinguent néanmoins des profils de notes dans le sens où ils représentent une estimation du contenu musical d'un extrait, et ne traduisent donc pas nécessairement les notes réellement jouées pendant cet extrait.

2.3 Conclusion

Nous avons présenté dans ce chapitre différentes informations tonales, en faisant apparaître un haut niveau de l'harmonie (accords et tonalités), correspondant à des abstractions musicales qui sont perçues par l'auditeur, et un bas niveau de l'harmonie (les notes de

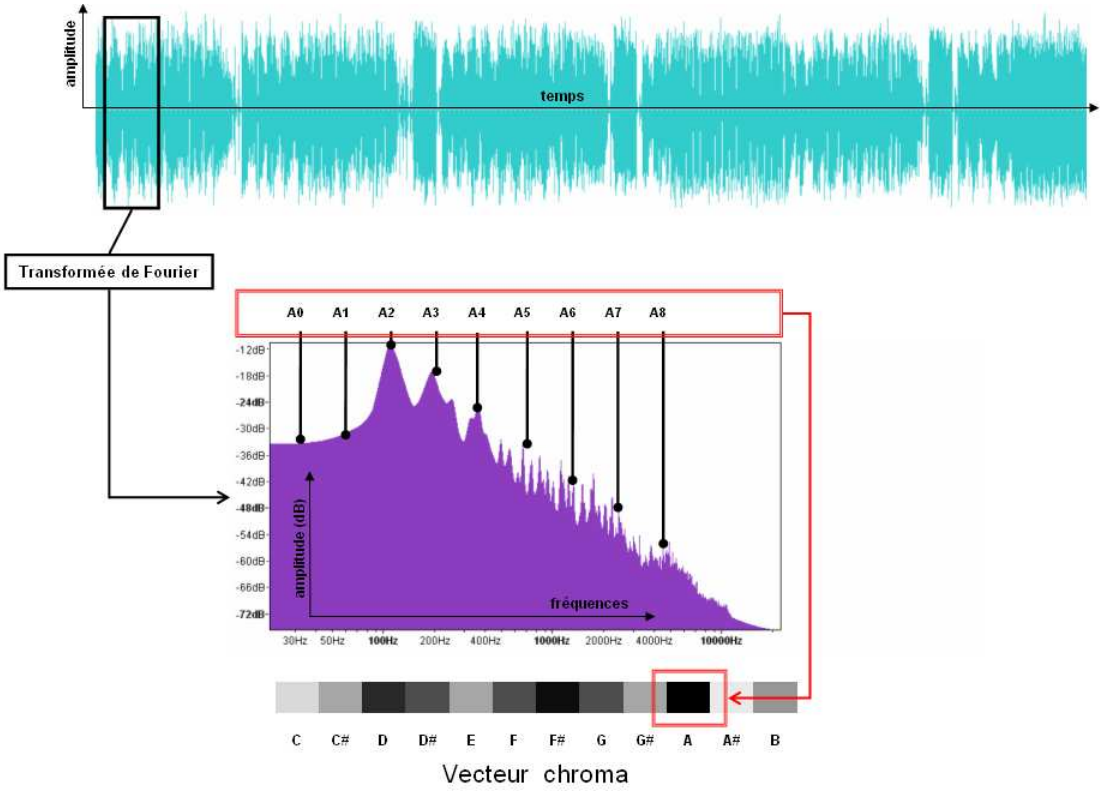


Figure 25 – Vue schématique de la construction d'un vecteur chroma à partir d'un spectre d'amplitude.

musique, qui sont concrètement jouées par les interprètes). Comme mentionné en introduction de ce chapitre, ces informations interagissent entre elles : les notes jouées participent aux accords, qui participent eux-mêmes à l'instauration d'une tonalité. Il est donc maintenant souhaitable d'étudier comment se modélisent et s'organisent ces interactions, selon quel type de hiérarchie et comment l'harmonie générale d'un morceau évolue avec chacune de ces informations.

Chapitre 3

Modèle de l'harmonie

Après avoir décrit les différentes informations tonales dans le chapitre précédent, nous étudions ici leurs interactions. Ces interactions peuvent exister entre deux informations de même nature, comme entre deux accords par exemple. On parlera dans ce cas de distance entre informations tonales. D'autres interactions existent entre informations de natures différentes, entre notes et tonalité par exemple. On parlera alors de structuration de l'harmonie. Nous étudions ainsi dans un premier temps les distances entre notes, puis entre accords (section 3.1), avant d'aborder différentes structurations de l'harmonie (section 3.2). Nous proposerons ensuite en section 3.3 un nouveau modèle hiérarchique permettant une organisation des différentes informations tonales.

3.1 Distances entre informations tonales

Nous étudions ici dans un premier temps les distances entre notes de musique, avant de présenter différentes distances entre accords. On notera que les distances musicales δ dont il est ici question ne correspondent pas à la définition mathématique de la distance, mais sont liées à la perception musicale. Ainsi, on observera en particulier dans de nombreux cas $\delta(x,y) \neq \delta(y,x)$.

3.1.1 Distances entre notes dans l'harmonie

La distance entre deux notes de musique peut être tout d'abord définie par la notion d'intervalle : on mesure le nombre de demi-tons séparant les deux notes, ou les deux classes de notes. Cette distance ainsi définie est bien liée à la perception : un Do suivi d'un Do# (1 demi-ton), est bien perçu comme engendrant un déplacement plus faible qu'un Do suivi d'un Ré (2 demi-tons). Néanmoins, au niveau de la consonance, la séquence (Do,Do#) apparaît plus dissonante que la séquence (Do,Ré). De ce point de vue, il semble préférable de considérer une autre distance, utilisant le cycle des quintes (représenté sur la figure 26).

Dans ce dernier, les classes de notes apparaissent le long d'un cercle, par intervalle de quinte juste (dans le sens horaire) ou de quarte juste (dans le sens anti-horaire). Il est ainsi possible de mesurer une distance entre deux classes de notes x et y en comptant le nombre de notes visitées pour aller de x à y en suivant le plus court chemin sur le cycle des quintes.

Ces deux distances entre notes ne dépendent pas du contexte tonal. Or, le passage d'un Do à un Sol en Do Majeur ne produit pas le même effet perceptif que cette même suite de notes en Sol Majeur. Il est donc souhaitable de pouvoir tenir compte de la tonalité pour la mesure de distance entre notes.

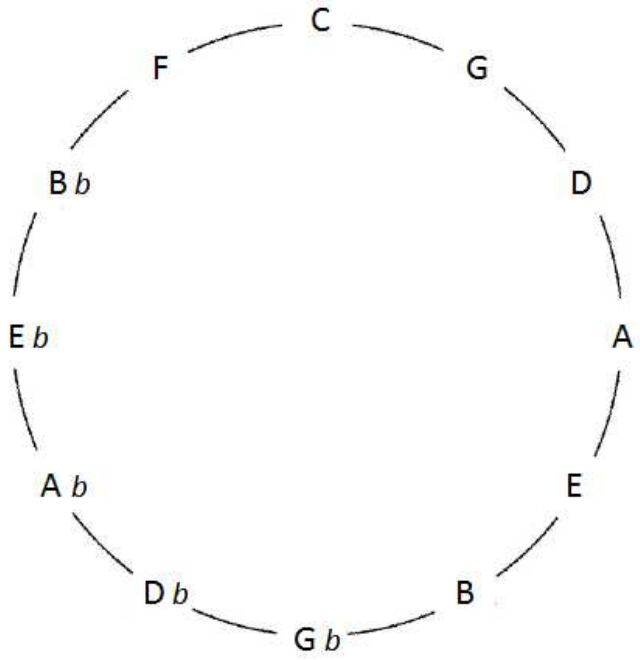


FIGURE 26 – Cycle des quintes présentant les classes de notes.

3.1.2 Profils de tonalité

Krumhansl s'est intéressée, en 1982, à la stabilité des classes de notes pour une tonalité donnée [KK82]. Une stabilité est définie comme étant "la mesure de l'appartenance d'une note à contexte tonal donné". Pour définir les valeurs de ces stabilités, Krumhansl a fait écouter à des sujets musiciens une suite de notes caractéristiques d'une tonalité, suivie d'une autre note dont la stabilité était à évaluer. Il était demandé aux sujets d'estimer numériquement la stabilité de cette dernière note selon la tonalité préalablement instaurée. La stabilité de chacune des classes de notes dans une tonalité donnée est représentée par une valeur dans un vecteur de dimension 12, de la même manière que dans un profil de notes. Deux vecteurs, l'un majeur et l'autre mineur, ont ainsi été définis (voir figure 27).

Ces vecteurs sont appelés profils de tonalité. Une fois un profil majeur (resp. mineur) construit, les profils des autres les tonalités majeures (resp. mineures) sont obtenus par rotation d'indice : Sol possède la même stabilité en Do Majeur que Mi en La Majeur par exemple. Ainsi, deux profils de tonalité décrivent la stabilité de chaque classe de note pour 24 tonalités différentes (12 tonalités majeures et 12 tonalités mineures).

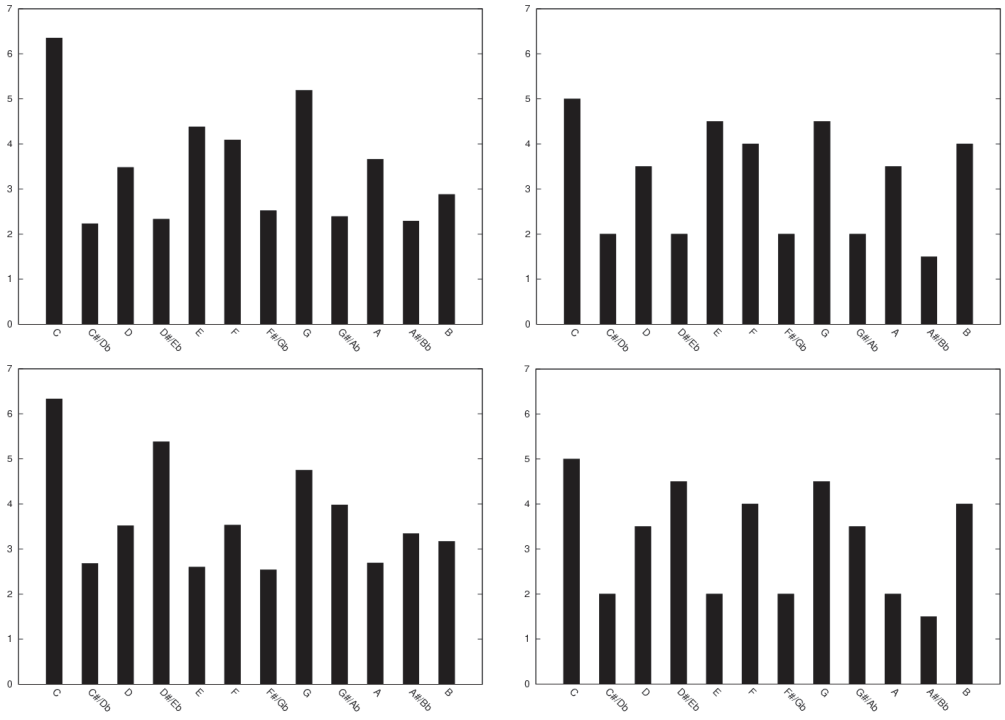


Figure 27 – Les profils de tonalité de Krumhansl (à gauche) et de Temperley (à droite) [Tem02], représentant les pondérations (en ordonnée) des différentes classes de notes (en abscisse). Le profil majeur est au dessus du profil mineur dans les deux cas.

En 2002, Temperley [Tem02] a apporté quelques modifications à ces profils (notamment dans la stabilité accordée à la sensible), en utilisant les résultats d'un algorithme d'estimation de tonalité. Les stabilités des classes de notes ont ainsi été modifiées pour maximiser la précision de cet algorithme. On trouvera les valeurs des profils de tonalité de Krumhansl et de Temperley dans le tableau 4, et une représentation de ces profils figure 27.

3.1.3 Mesure entre accords dans l'harmonie

La musique tonale étant caractérisée, harmoniquement parlant, de progressions d'accords, la mesure musicale entre deux accords successifs constitue un enjeu majeur du processus de composition ou d'arrangement. L'étude des cadences, ces séquences de deux

profil	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
K-maj	6.35	2.23	3.48	2.33	4.38	4.09	2.52	5.19	2.39	3.66	2.29	2.88
K-min	6.33	2.68	3.52	5.38	2.60	3.53	2.54	4.75	3.98	2.69	3.34	3.17
T-maj	5.0	2.0	3.5	2.0	4.5	4.0	2.0	4.5	2.0	3.5	1.5	4.0
T-min	5.0	2.0	3.5	4.4	2.0	4.0	2.0	4.5	3.5	2.0	1.5	4.0

TABLE 4 – Les profils de tonalités de Krumhansl [KK82] (K-maj pour le profil de Do Majeur et K-min pour le profil de Do mineur), et de Temperley [Tem02] (T-maj pour le profil de Do Majeur et T-min pour le profil de Do mineur).

C	C#	D	D#	E	F	F#	G	G#	A	A#	B
1	1	0	0	0	1	0	1	0	0	0	1

TABLE 5 – Vecteur de potentiel d'affinité pour la classe de Do.

ou trois accords permettant selon les cas d'affirmer une tonalité, ou au contraire de permettre une possible modulation, renforce en particulier l'intérêt d'une mesure musicale entre accords. Cette section présente ainsi différentes propositions de distances musicales entre accords.

3.1.3.1 Affinités entre accords

Costère, en 1962 [Cos62], définit le potentiel d'affinité des différentes classes de notes. Ce potentiel est représenté, pour chaque classe, par un vecteur de dimension 12, comportant les valeurs d'affinité avec les autres classes. Dans le vecteur de la classe de note x, les valeurs sont nulles à l'exception des valeurs correspondant à x et à x augmenté de un, cinq, sept et onze demi-tons. On trouvera le vecteur potentiel d'affinité pour la classe de Do dans le tableau 5.

Le vecteur potentiel d'affinité d'un accord s'obtient en sommant les vecteurs correspondant aux classes des notes le composant. On trouvera le vecteur potentiel d'affinité de Do

	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
vect. C	1	1	0	0	0	1	0	1	0	0	0	1
vect. E	0	0	0	1	1	1	0	0	0	1	0	1
vect. G	1	0	1	0	0	0	1	1	1	0	0	0
vect. $CMaj$	2	1	1	1	1	2	1	2	1	1	0	2

TABLE 6 – Vecteur de potentiel d'affinité pour l'accord de Do Majeur, obtenu en sommant les vecteurs d'affinité des notes de la triade de Do Majeur : Do, Mi et Sol.

Majeur dans le tableau 6. L'affinité entre deux accords est ensuite estimée selon la formule suivante :

$$A(C_1, C_2) = \sum_{i=1}^{\#(C_2)} V_{C_1}(C_2(i)) \tag{1}$$

où C_1 et C_2 sont deux accords, $\#(C_2)$ le nombres de notes différentes dans C_2, V_{C_1} le vecteur potentiel d'affinité de C_1 et $C_2(i)$ la $i$ème note de C_2.

Ainsi,

$$\begin{aligned} A(G_M, D_M) &= \sum_{i=1}^{3} V_{C_1}(C_2(i)) \\ &= V_{C_1}(2) + V_{C_1}(6) + V_{C_1}(9) \\ &= 2 + 2 + 1 = 5 \end{aligned}$$

Il est possible de proposer une mesure de distance entre accords d'après cette définition de l'affinité. Cette distance, dont les valeurs s'étalent entre 0 (aucune affinité) et 1, peut être définie comme suit :

$$\delta(C_1, C_2) = 1 - \frac{A_{max}(C_1) - A(C_1, C_2)}{A_{max}(C_1)} \tag{2}$$

avec

$$A_{max}(C_1) = \sum_{i=1}^{12} V_{C_1}(i) \tag{3}$$

3.1.3.2 Contenu d'intervalle

Dans la théorie musicale des ensembles proposée en 1960 par Hanson [Han60], un vecteur décrit les intervalles séparant les différentes notes d'un même ensemble. Un tel vecteur se présente sous la forme d'une suite de six chiffres, qui désignent, dans l'ordre, le nombre de :

- secondes diminuées (= 1 demi-ton),
- secondes justes (= 2 demi-tons),
- secondes augmentées ou tierces mineures (= 3 demi-tons),
- tierces majeures (= 4 demi-tons),
- quartes justes (= 5 demi-tons),
- quartes augmentées ou quintes diminuées (= 6 demi-tons).

Ce vecteur fut nommé vecteur de contenu d'intervalles (*intervallic content*) par Hanson et a ensuite été repris par Forte [For73].

A titre d'exemple, le vecteur d'intervalles de l'ensemble de notes [Do Mi Sol] est :

$$< 001110 >$$

Cet ensemble comporte en effet une tierce mineure (de Mi à Sol), une tierce majeure (de Do à Mi) et une quarte juste (de Sol à Do). Il est à noter qu'un vecteur d'intervalles est invariant par transposition de l'ensemble des notes, de même que par échange entre notes.

En utilisant cette notion, une distance entre accords utilisant les vecteurs d'intervalles a été proposée en 2002 [Mat02] en utilisant le formalisme suivant :

$$\delta(C_1, C_2) = \sum_{i=1}^{6} |V_{C_1}(i) - V_{C_2}(i)| \tag{4}$$

où C_1 et C_2 sont deux accords, et V_{C_1} et V_{C_2} leurs vecteurs d'intervalle respectifs. Une telle distance peut prendre des valeurs comprises entre 0 et 66.

3.1.3.3 Distance d'Estrada

La distance d'Estrada [Tru04, Mat02], proposée par le compositeur Bloch avec l'aide d'Estrada, s'appuie sur une comparaison des intervalles entrant en jeu dans deux accords successifs, en utilisant la notion de *texture*. Une *texture* est obtenue en ordonnant, selon leurs hauteurs, les différentes notes formant un accord au sein d'une octave formée par la plus grave de ces notes et sa répétition à l'octave supérieure. Aucun renversement d'accord n'est ainsi pris en compte. Les différents intervalles entrant en jeu dans l'organisation de la texture sont ensuite décrits par un vecteur de contenu d'intervalle (voir section 3.1.3.2). La distance entre deux accords peut alors être exprimée par :

$$\delta(C_1, C_2) = max(\#(C_1), \#(C_2)) - |V_{C_1} \cap V_{C_2}| - 1 \tag{5}$$

où C_1 et C_2 sont deux accords, $\#(C_1)$ et $\#(C_2)$ leurs nombres de notes respectifs et V_{C_1} et V_{C_2} leurs vecteurs de contenu d'intervalle respectifs. Si le résultat est négatif, la distance est ramenée à 0. Le maximum théorique de cette distance est 11.

3.1.3.4 Distances géométriques

Dans la continuité du *Tonnetz* proposé au XVIIIème siècle par Euler [Eul39], plusieurs distances géométriques ont été proposées afin de visualiser l'évolution de l'harmonie dans la musique. En 2000, Chew [Che00] introduit notamment un modèle mathématique de l'harmonie en spirale (voir figure 28). Les classes de notes sont situées sur cette spirale de la manière suivante : un déplacement d'un quart de tour correspond à un intervalle de quinte, alors qu'un tour entier correspond à une tierce majeure. Le rayon r du cylindre dans lequel est inscrit la spirale et le gain de hauteur par quart de tour h définissent un

positionnement unique des classes de notes le long de cette spirale, pouvant être décrit par :

$$P(k) = \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix} = \begin{bmatrix} r\sin(\frac{k\pi}{2}) \\ r\cos(\frac{k\pi}{2}) \\ kh \end{bmatrix} \tag{6}$$

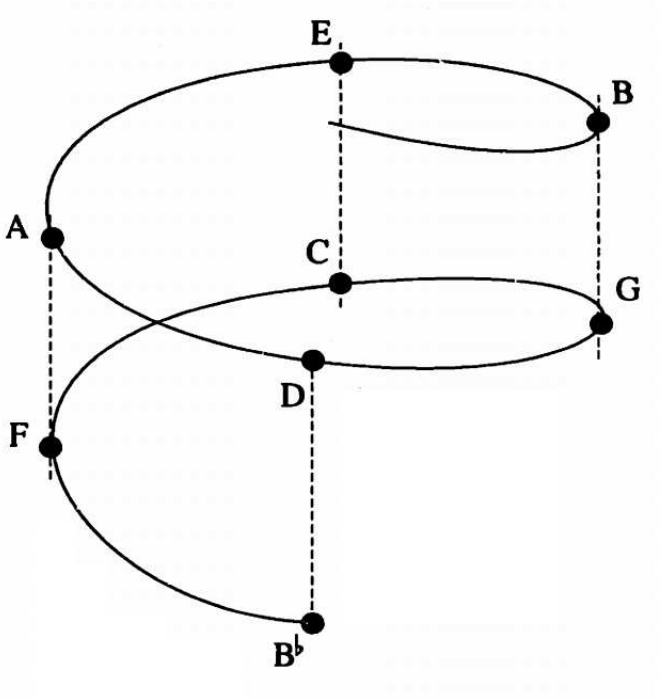


Figure 28 – Modèle en spirale d'après [Che00].

Dans ce modèle en spirale, un accord a peut être représenté par le centre de gravité du triangle représentant les classes de notes de la fondamentale, la tierce et la quinte de a :

$$C_M(k) = w_1 \cdot P(k) + w_2 \cdot P(k+1) + w_3 \cdot P(k+4) \tag{7}$$

où les poids w_1,w_2 et w_3 vérifient $w_1 \geq w_2 \geq w_3 > 0$ et sont tels que : $\sum_{i=1}^{3} w_i = 1$.

Nous proposons d'estimer une distance entre accords suivant ce modèle, en calculant la distance géométrique séparant, dans la spirale, deux points représentant les deux accords. On supposera que les classes de notes composant ces accords sont telles que leurs poids respectifs sont égaux. Les valeurs possibles pour cette distance sont ainsi contenues dans l'intervalle $[0; +\infty[$.

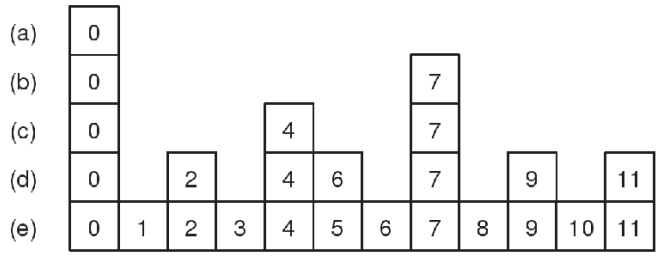


FIGURE 29 – Espace basique de l'accord de Do Majeur dans la tonalité de Do Majeur.

3.1.3.5 Espace tonal (*Tonal Pitch Space*)

Lerdahl définit en 2001 [Ler01] l'espace basique (*basic space*) d'un accord a dans une région tonale t (les notions de régions tonales et de tonalités peuvent être confondue [RK10]) comme la superposition géométrique :

- des notes chromatiques de la tonalité t (niveau chromatique),
- des notes diatoniques de la tonalité t (niveau diatonique),
- des notes de la triade de l'accord a (niveau triadique),
- de la fondamentale et la quinte de l'accord a (niveau des quintes),
- de la fondamentale de l'accord a (niveau fondamental).

La figure 29 montre l'espace basique de l'accord de Do Majeur dans la tonalité de Do Majeur.

Si (C_x, K_x) représente l'accord C_x dans la tonalité K_x, Lerdahl définit la distance entre $x = (C_x, K_x)$ et $y = (C_y, K_y)$ comme étant la somme de trois termes :

$$\delta(x, y) = i + j + k \tag{8}$$

i et j sont respectivement les distances entre K_x et K_y et entre C_x et C_y dans le cercle des quintes. Cette distance sur le cercle des quintes est obtenue en dénombrant le nombre minimal d'accords (ou de tonalités) à visiter sur le cercle pour passer de C_x à C_y (ou de K_x à K_y). Ainsi, la distance entre Do Majeur et Sol Majeur sur ce cercle est 1. Il en va de même pour la distance entre Do Majeur et Fa Majeur. La distance entre Fa Majeur et Sol Majeur est de 2.

k est le nombre de notes apparaissant dans l'espace basique de y et pas dans l'espace basique de x, ce qui revient à superposer géométriquement les deux espaces basiques et à dénombrer les notes apparaissant dans celui de y et pas dans celui de x.

Un exemple de calcul de distance est proposé figure 30. Dans cet exemple, $x = (C_M, C_M)$

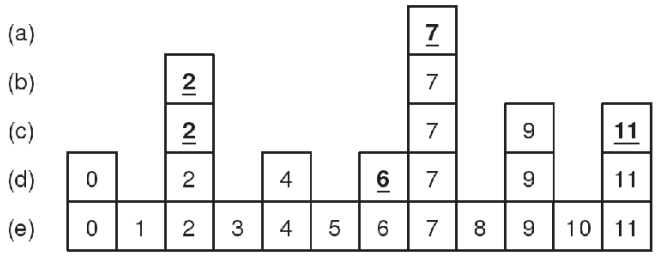


Figure 30 – Illustration de la distance entre $x = (C_M, C_M)$ et $y = (G_M, G_M)$.

et $y = (G_M, G_M)$. Ici, i=j=1 car la distance entre Do Majeur et Sol Majeur sur le cercle des quintes est 1, que ce soit pour les accords ou pour les tonalités. k=5 car il y a 5 notes apparaissant dans l'espace basique de y et non dans l'espace de x. Ces notes sont soulignées dans la figure 30. La distance entre $x = (C_M, C_M)$ et $y = (G_M, G_M)$ est donc égale à 7. On notera que les valeurs possibles pour la distance de Lerdahl sont comprises entre 0 et 14.

Cette distance n'est pas, ainsi définie, symétrique (on a $\delta(x, y) \neq \delta(y, x)$). Il est cependant possible de symétriser cette distance, en s'inspirant de la modification apportée à la distance de Lerdahl dans [DHVW08]. Pour ce faire, nous choisissons de définir L_{sym} (distance de Lerdahl symétrique) en fonction de L (distance de Lerdahl asymétrique) de la manière suivante :

$$L_{sym}(x, y) = \frac{L(x, y) + L(y, x)}{2} = L_{sym}(y, x) \tag{9}$$

3.1.3.6 Distance de Paiement

Paiement propose en 2005 une nouvelle distance entre accords [PEB05]. Il suppose que pour chaque accord C_j, la représentation distribuée $l_j = l_j(n_1), ..., l_j(n_d)$ correspond à l'intensité perçue des harmoniques en fonction des classes de notes n_k, seules les d premières harmoniques étant perceptibles. Il propose ainsi de considérer la fonction m permettant d'associer un *pitch* MIDI à une fréquence f :

$$m(f) = 12 \log_2(f) - \log_2(8,1758)) \tag{10}$$

et propose la définition suivante pour $l_j(n)$:

$$l_j(n) = \max_{h \in \mathbb{N}, i \in C_j} (\{\rho^h | \text{round}(m(f_{h,i})) = n\} \cup \{0\}) \tag{11}$$

	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
Majeur	0	100	70	97	48	55	94	27	97	67	100	85
Mineur	0	96	73	27	96	73	100	42	62	96	77	85

TABLE 7 – Valeurs d'instabilités normalisés pour la tonalité de Do Majeur et Do mineur

unisson	2min	2maj	3min	3maj	4	triton	5	6min	6maj	7min	7maj
0	91	64	46	37	18	100	9	55	36	73	82

TABLE 8 – Valeurs d'instabilités de chaque intervalle.

La valeur choisie pour ρ est 0.97. Il définit ensuite un vecteur à 12 dimensions v_j mesurant l'intensité relative de chaque classe de notes pour un accord C_j. Ici, $v_j(0)$ représente la classe Do, $v_j(1)$ la classe Do#, etc..

$$v_j(i) = \sum_{n_k:1\leq k\leq d,(n_k mod 12)=i} l(n_k) \tag{12}$$

D'après cette représentation, nous nous proposons de définir une distance euclidienne entre deux accords, avec des valeurs situées dans l'intervalle $[0; +\infty[$.

3.1.3.7 Distance de *Pitch-Interval*

Min-Joon Yoo et In-Kwon Lee [YL06], en 2006, proposent d'utiliser les travaux de Krumhansl et Shepard [KS79] pour estimer une distance musicale entre accords. La notion de stabilité tonale de certaines classes de notes par rapport à une tonalité donnée permet ainsi de créer des valeurs d'instabilité normalisées, comme présentées dans le tableau 7.

Néanmoins, Min-Joon Yoo et In-Kwon Lee mentionnent que la tension musicale entre accords résulte non seulement des classes de notes, mais également des intervalles entre notes à l'intérieur même des accords. Ainsi, ils utilisent également des valeurs d'instabilité pour les différents intervalles, présentées dans le tableau 8.

Dans ce modèle, le degré de tension d'un accord est calculé comme suit :

$$d(C) = w_1 \sum s + w_2 \sum r \tag{13}$$

où s est l'instabilité des classes de notes composant l'accord C, r est l'instabilité des intervalles de cet accord C, et $w1$ et $w2$ sont des pondérations pour ces deux termes. Les auteurs précisent que $w1$ et $w2$ sont affectés, à l'implémentation, à des valeurs qui rendent équivalentes l'influence des deux facteurs. Les valeurs prises par cette distance dépendent des poids et peuvent théoriquement s'étaler dans l'intervalle $[0 : +\infty[$.

3.1.4 Propriétés des différentes distances et exemples

Le tableau 9 présente les propriétés de chaque distance entre accords présentée précédemment, ainsi que leurs valeurs pour quelques enchaînements d'accords types. CM_{GM} représente la triade de Do Majeur dans la tonalité de Sol Majeur. *Cdiat* désigne un accord composé des 12 classes de notes. Une distance entre fondamentales est également présentée dans ce tableau, et mesure simplement la distance entre les fondamentales des deux accords sur le cycle des quintes. On observe que toutes les distances vérifient $\delta(x, x) = 0$, sauf la distance de Costère. De plus, cette distance présente des valeurs identiques pour toutes les transitions entre accords de triades majeurs, et n'est donc pas très discriminante lorsque seuls les accords de triades sont considérés, à l'inverse des distances de Paiement et de *Pitch-Interval* qui semblent induire les différences les plus importantes pour les transitions présentées. Les distances de Lerdahl et de *Pitch-Interval* mises à part, les tonalités n'interviennent pas dans le calcul de distance entre deux accords. On notera que pour la distance de Lerdahl, le changement de tonalité augmente automatiquement la valeur mesurée, alors que la distance de *Pitch-Interval* utilise le caractère relatif de l'accord à la tonalité associée, ce qui explique par exemple que $\delta(CM_{CM} \rightarrow GM_{GM}) = 0$.

3.2 Représentations hiérarchiques

Selon C.S. Lee [Lee85], la perception de quasiment n'importe quelle pièce musicale ne relève non pas d'une séquence arbitraire de notes, mais de structures temporelles dans lesquelles les notes sont regroupées en différents types d'objets. Les représentations hiérarchiques de l'harmonie permettent ainsi de grouper entre eux les objets élémentaires de l'harmonie, les notes, pour obtenir des objets parents, dérivant eux-mêmes d'autre objets, etc. La figure 31 illustre un exemple de regroupement hiérarchique des notes, laissant apparaître d'autres objets suggérant une modélisation hiérarchique de la musique. La plupart des représentations hiérarchiques de la musique concernent le rythme et la mélodie [Lee85, LH78, Mar72, SL92], ou même la structure [Lin92], voir figure 32. Toutefois, quelques autres travaux, présentés dans cette section, se sont penchés sur de possibles représentations hiérarchiques de l'harmonie.

3.2.1 Notions sur les arbres

La formalisation d'un modèle hiérarchique se fait souvent à l'aide d'un objet mathématique appelé *arbre*. Un arbre est un ensemble de *sommets*, reliés entre eux par des *arêtes* et pouvant se représenter sous la forme d'une hiérarchie. Chaque sommet peut présenter un ou plusieurs *fils*, et possède un unique *père*, à l'exception de la *racine*, qui ne possède pas de père. Dans le cas d'un sommet sans fils, on parle de *feuille*. Le *niveau* d'un sommet

	Lerdahl	Spirale	Estrada	Contenu-Int.	Paiement	Costère	Lerdahl$_{sym}$	Pitch-Int.	Fondamentales
type	entier	réel	entier	entier	réel	réel	réel	réel	entier
intervalle	$[0:14]$	$[0:+\infty[$	$[0:11]$	$[0:66]$	$[0:+\infty[$	$[0:1]$	$[0:14]$	$[0:+\infty]$	$[0:6]$
$CM_{CM} \rightarrow CM_{CM}$	0	0.0	0	0	0.0	0.33	0.0	0.0	0
$CM_{CM} \rightarrow GM_{CM}$	5	0.49	0	0	2.28	0.33	5.0	1.07	1
$FM_{CM} \rightarrow CM_{CM}$	5	0.57	0	0	2.35	0.33	5.0	0.47	1
$CM_{CM} \rightarrow Am_{CM}$	7	1.25	0	0	1.49	0.27	7.0	0.40	3
$CM_{CM} \rightarrow CM_{GM}$	2	0.0	0	0	0.0	0.33	2.0	0.47	0
$CM_{CM} \rightarrow GM_{GM}$	7	0.49	0	0	2.28	0.33	7.0	0.0	1
$CM_{CM} \rightarrow CM^7_{CM}$	2	0.28	1	3	0.83	0.33	1.0	3.19	0
$CM_{CM} \rightarrow GM^7_{CM}$	6	0.54	1	3	2.46	0.47	5.5	3.81	1
$CM_{CM} \rightarrow Cdiat_{CM}$	14	0.60	11	63	4.08	1.00	14.0	43.51	0

Table 9 – Mesures d'enchaînements d'accords en fonction des différentes distances musicales considérées.

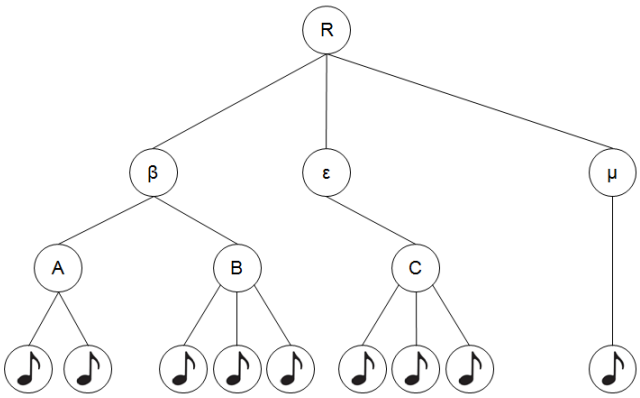


Figure 31 – Exemple de regroupement des notes suivant une représentation hiérarchique. Les notes sont regroupées en objets (A,B,C) qui sont ensuite eux-mêmes regroupés en autres objets (β, ϵ, μ). La racine est représentée par le nœud R.

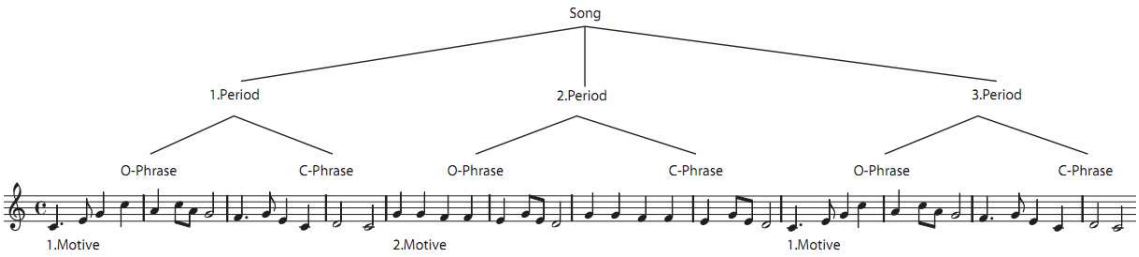


Figure 32 – Exemple de représentation hiérarchique de la structure, d'après [Lin92].

est un entier représentant le nombre de pères successifs qu'il faut remonter pour parvenir à la racine. La *profondeur* d'un arbre est le niveau maximal présenté par (au moins) une de ses feuilles.
Sur la figure 31, les notes sont ainsi les feuilles de l'arbre.

3.2.2 Analyse schenkerienne

Schenker [Sch35] propose au début du XXème siècle une méthode d'analyse musicale composée de transformations successives, appelées réductions. Ces réductions visent, pour chaque motif mélodique élémentaire, à déterminer une note primitive représentant ce motif. Les réductions sont successives, et aboutissent finalement à une seule note, représentant la tonique de la tonalité associée à la tonalité de la pièce. Il est ainsi possible de représenter les réductions successives sous forme hiérarchique, illustrant ainsi l'arborescence des élaborations constituant la pièce étudiée. La figure 33 montre un exemple de réduction schenkerienne : la figure du haut montre la réduction sur la première phrase de la sonate

pour piano en La de Mozart (K.331), et la figure du bas montre une représentation de la réduction sous forme hiérarchique, avec les règles appliquées pour parvenir à cette réduction. Dans cette figure, les notes réelles de la pièce sont surlignées en rouge. Les autres notes sont des réductions (et par conséquent des notes non réellement jouées), issues des règles respectives encadrées.

Néanmoins, il n'existe pour l'heure aucune implémentation de la réduction d'une pièce selon l'analyse schenkerienne. Plusieurs travaux se sont pourtant penchés sur une possible automatisation de cette analyse : Smoliar [Smo79], Kassler [Kas88], Kirlin and Utgoff [KU08], et Marsden [Mar05, Mar07]. Mais dans tous les cas, une multitude de situations se produisent où un choix doit être fait entre plusieurs réductions différentes, issues de plusieurs règles concurrentes. Il n'existe donc pas de manière unique de parvenir à une réduction, et donc à une représentation hiérarchique de l'harmonie, d'après l'analyse schenkerienne.

3.2.3 Théorie Génératrice de la Musique Tonale

En 1983, Lerdahl et Jackendoff proposent une autre méthode d'analyse dans leur Théorie Génératrice de la Musique Tonale [LJ85] (*Generative Theory of Tonal Music, ou simplement GTTM*). Cette méthode est similaire à l'analyse schenkerienne, et utilise une série de règles pour regrouper deux notes successives, et leur choisir une unique note représentante. Ces règles utilisent des informations relatives au tempo, à la structure et aux relations harmoniques et mélodiques entre différentes notes. Lerdahl et Jackendoff utilisent le terme de "subordination" pour qualifier la note représentant un groupe de deux notes consécutives ; ce terme pouvant se rapprocher de la réduction définie par Schenker. S'il s'agit de la note de gauche, on parle d'élaboration à droite, et vice versa pour la note de droite. La figure 34 présente une représentation hiérarchique d'élaboration à droite, où la note de droite d constitue la note la plus prépondérante, et représente d et g au niveau hiérarchique supérieur. Tout comme dans l'analyse schenkerienne, les élaborations sont successives, et forment une arborescence de transformations. La figure 35 offre une représentation de l'analyse par Lerdahl et Jackendoff du début de la sonate pour piano en La de Mozart (K.331).

La théorie génératrice de la musique tonale comporte 4 grandes familles de règles :

- les règles de regroupement (*grouping analysis rules*),
- les règles de métrique (*metrical analysis rules*),
- les règles de regroupement en arbre temporel (*time span tree rules*),
- les règles de rallongement (*prolongational tree rules*).

Les trois premières familles concernent l'analyse rythmique, et ne peuvent donc pas être utilisées directement à des fins harmoniques. Satoshi Tojo, co-auteur d'un tutoriel sur

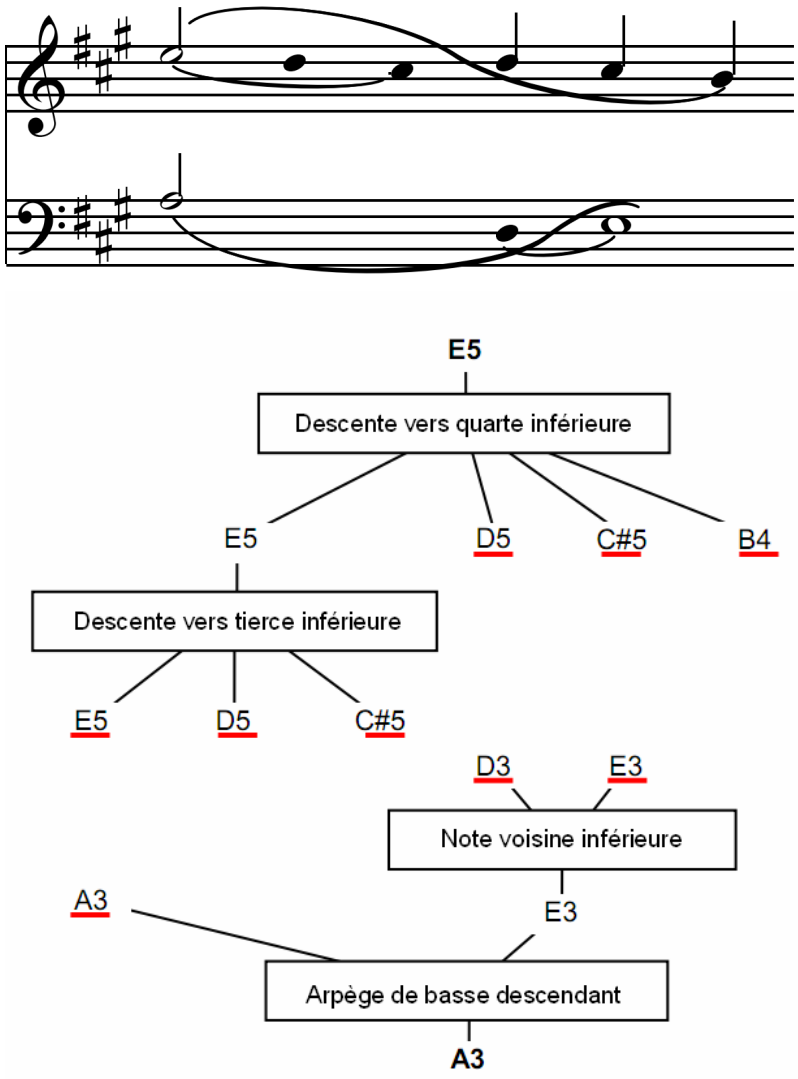


Figure 33 – Exemple de réduction schenkerienne (d'après [Sch35]), et de représentation hiérarchique de cette réduction. Les notes soulignées correspondent aux notes réellement jouées, les autres étant des réductions.

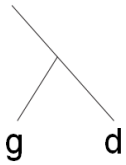


FIGURE 34 – Exemple de représentation d'élaboration à droite entre deux notes g et d, d'après la Théorie Génératrice de la Musique Tonale.

les règles GTTM [HTH07] a non seulement confirmé [1] l'impossibilité d'utiliser les trois premiers groupements de règles en vue d'une analyse de l'harmonie, mais a également affirmé que le quatrième regroupement ne permettait pas d'obtenir une hiérarchisation harmonique sans apport d'informations extérieures à la partition musicale (comme le degré d'un accord, par exemple). En d'autres termes, les règles d'analyses GTTM ne permettent pas d'obtenir certaines informations tonales dont elles ont besoin pour procéder aux subordinations, ces informations devant alors être annotées à la main.

Comme c'est le cas pour l'analyse schenkerienne, la Théorie Génératrice de la Musique Tonale ne peut pas être formellement implémentée en suivant les seuls travaux de Lerdahl et Jackendoff. D'autres travaux [HTH07, JSM93] se sont penchés sur de possibles implémentations de cette théorie, en superposant un autre système de règles pour pouvoir traiter les cas de conflits entre les règles pouvant s'appliquer simultanément. Dans ces travaux, l'utilisateur doit néanmoins toujours intervenir, et aucune analyse non supervisée d'une pièce de musique entière n'est encore envisageable.

La théorie génératrice de la musique tonale présente en outre les même propriétés que l'analyse schenkerienne en terme de profondeur d'arbre hiérarchique variable, ou en terme de coexistence d'objets musicaux. Sur la figure 35, on observe notamment que le niveau d fait coexister des subordinations de notes avec des notes réellement jouées.

3.2.4 Le modèle cognitif de Bharucha

En 1987, Bharucha [Bha93] proposa une représentation tonale mettant en jeu trois niveaux : notes, accords, et tonalités (voir figure 36). L'idée principale de la représentation est la suivante : chaque note jouée est successivement activée pendant une courte durée, et active par proximité les accords auxquels elle est reliée. Une note est reliée à tous les accords dont elle compose la triade. Do est ainsi relié à l'accord de Do Majeur, Fa Majeur, La♭ Majeur, mais aussi La Mineur, Fa Mineur et Do Mineur. Si d'autres notes viennent à être jouées alors qu'une note est toujours active, elles peuvent renforcer l'activation de certains accords (par exemple, si Do, Mi et Sol sont joués de sorte que leurs durées d'activité

1. contact personnel avec l'auteur
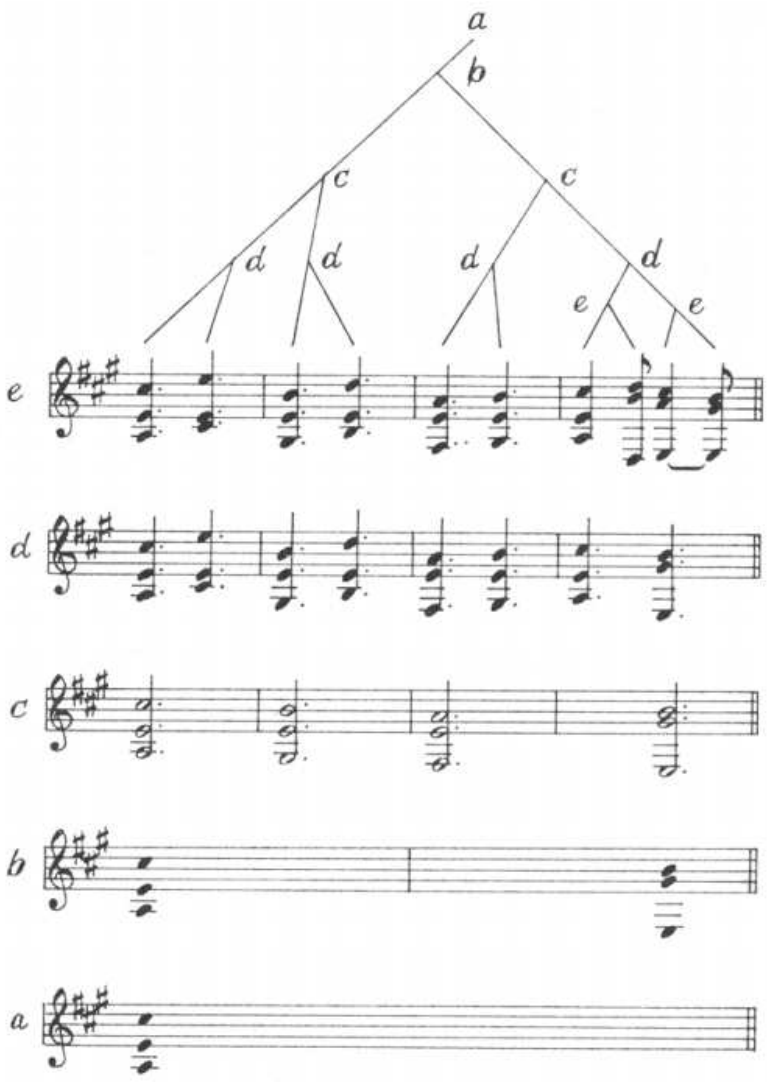


Figure 35 – Analyse par Lerdahl et Jackendoff du début de la sonate pour piano en La de Mozart (d'après [LJ85]). Chaque sommet de l'arbre de réduction est annoté à l'aide d'une lettre, qui désigne le niveau auquel la réduction associée prend place. Les différents niveaux sont représentés en dessous de l'arbre.

interfèrent, alors l'accord de Do Majeur est l'accord qui se trouve le plus fortement activé). Le même principe s'applique entre les niveaux *accords* et *tonalités*. Sur la figure 36, chaque note (zone verte) est reliée aux trois accords de triades majeures (zone bleu foncé) et aux trois accords de triades mineures (zone bleu clair) qui la contiennent. Chaque accord est ensuite relié aux trois tonalités (zone rouge) l'acceptant comme Ier, IVème et Vème degré.

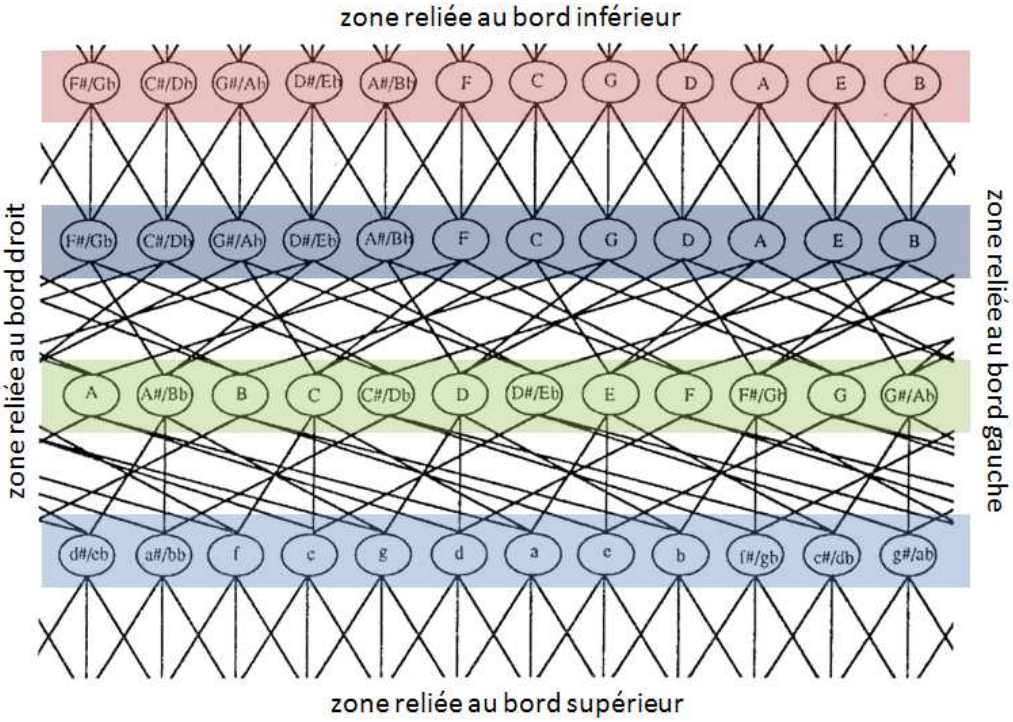


FIGURE 36 – La représentation tonale de Bharucha [Bha93], qui est une surface de tore : les extrémités droites et gauches sont ainsi reliées, comme les extrémités hautes et basses.

La représentation de Bharucha présente donc plusieurs niveaux organisés sur la surface d'un tore. Cette représentation ne représente pas un arbre, mais un graphe, car si plusieurs niveaux hiérarchiques sont présents, aucune racine n'existe dans cette représentation. De plus, une certaine confusion demeure sur ce que Bharucha considère comme un accord : il semble en effet difficile de savoir si cette définition rejoint notre définition d'un groupe de notes, ou plutôt celle d'un accord. Selon lui, la durée d'activation tend à signifier que l'accord peut perdurer après que les notes qui le composent aient fini de sonner (ce qui pourrait donc rejoindre notre définition d'un accord). En revanche, les dates de début des notes prises en compte dans l'activation des accords influent sur la nature de ces derniers, qui seraient donc plus proches de notre définition du groupe de notes (puisque les notes doivent sonner en même temps, ou dans le même temps d'activation, pour faire partie d'un même accord selon Bharucha). La même ambiguïté existe pour la tonalité : telle qu'elle est décrite, la représentation de Bharucha suppose qu'un changement de tonalité est possible,

conférant ainsi aux tonalités détectées un caractère local. Il semble donc difficile de pouvoir définir une tonalité principale à l'aide de cette représentation.

3.2.5 Le modèle hiérarchique de Rizo

En 2006, Rizo [RIPdL06] propose un modèle hiérarchique de la musique polyphonique symbolique utilisant des règles de groupement rythmique entre notes. Ce modèle présente également des règles pour l'estimation de tonalité générale. Le fonctionnement est le suivant :

- les notes sont groupées selon la hiérarchie rythmique présentée figure 37,
- chaque nœud est étiqueté par les classes de notes rentrant en composition dans la descendance de ce nœud, comme illustré figure 38. Ainsi, un nœud ayant deux fils sera étiqueté par l'union des classes de notes labellisant ses deux fils,
- chaque nœud fait l'objet d'un classement de tonalités candidates, en fonction des classes de notes labellisant ce nœud,
- les classements des tonalités candidates pour chaque nœud sont considérés, et un algorithme récursif permet d'établir le classement final des tonalités candidates, correspondant à la racine de l'arbre.

Comme la réduction schenkerienne, cette approche utilise une modélisation hiérarchique de l'harmonie par un arbre de profondeur variable. Ce modèle se différencie néanmoins des travaux précédemment présentés dans le sens où celui-ci est formalisé et implémenté. Des expériences sur 212 fichiers MIDI ont été menées et montrent que l'estimation de la tonalité en utilisant ce modèle est légèrement plus précise que pour les méthodes implémentées par le logiciel *Melisma* [TS], et que la complexité de cette méthode leur est inférieure en requérant un temps de calcul près de 4 fois plus petit.

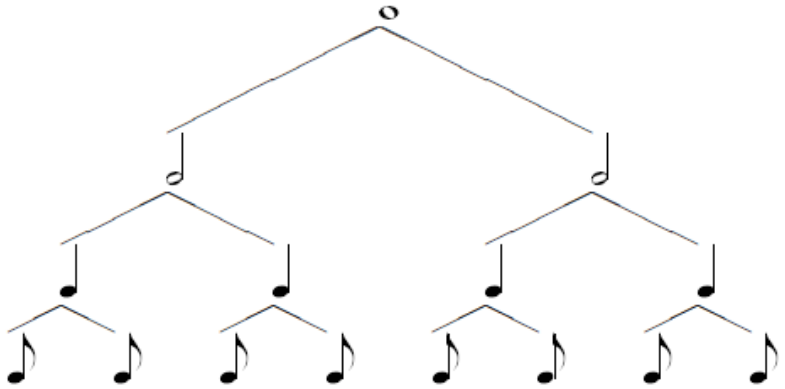


Figure 37 – Hiérarchie rythmique pour le groupement de notes, d'après [RIPdL06].

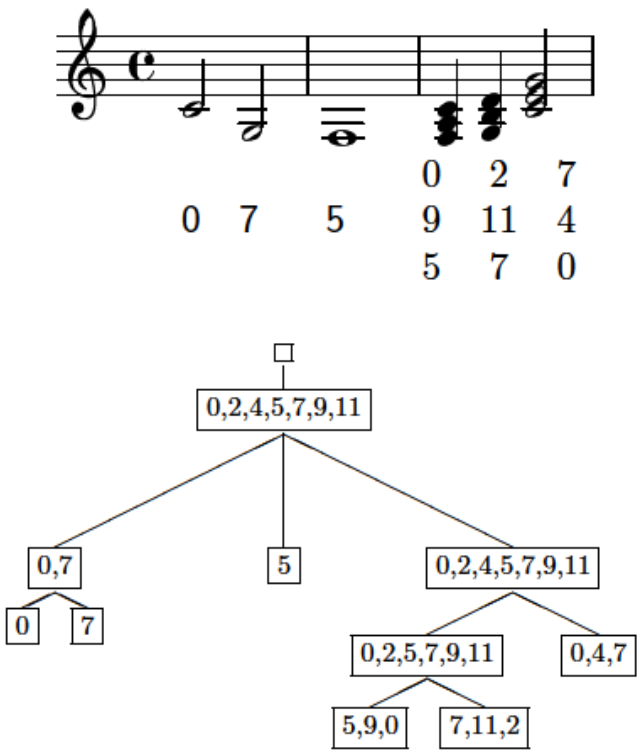


FIGURE 38 – Exemple de détection de tonalité d'après le modèle hiérarchique proposé par Rizo, d'après [RIPdL06].

Parmi les modèles hiérarchiques présentés, les modèles de Schenker et de Lerdahl (*GTTM*) ne sont pas directement implémentables. Le modèle de Rizo utilise des règles de groupement rythmique, qui aboutissent à des représentations différentes d'une même mélodie si la signature rythmique est différente, par exemple. En outre, il n'est pas possible d'accéder directement à une information tonale donnée avec ce modèle, dans lequel la séquence d'accords et de tonalités locales n'apparaissent pas.

3.3 Arbre d'harmonie

Nous proposons ici un nouveau modèle hiérarchique de l'harmonie. Une description de ce modèle sera d'abord présentée, avant que ces propriétés ne soient abordées.

3.3.1 Description

Nous nous proposons de représenter l'harmonie d'un morceau de musique à l'aide d'un arbre ordonné temporellement (voir figure 39) : l'ordre des nœuds frères est déterminé par la date d'apparition des propriétés tonales concernées. Le temps apparaît ainsi de gauche à

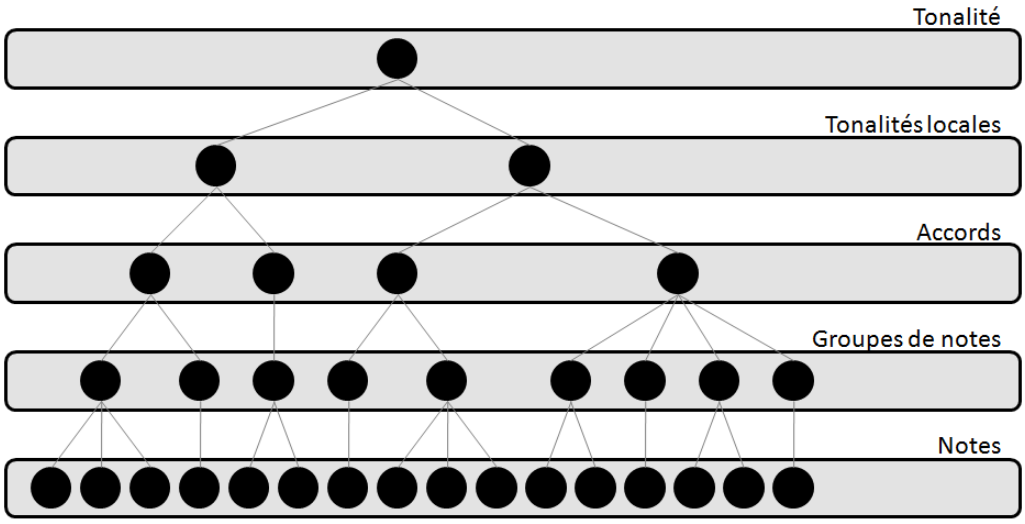


Figure 39 – Arbre d'harmonie proposé, composé de cinq niveaux représentant cinq informations tonales différentes.

droite sur chaque niveau. L'arbre proposé comporte cinq niveaux distincts, chacun de ces niveau faisant référence aux propriétés tonales décrites au chapitre 2. Ainsi :

- la tonalité générale du morceau de musique est représentée par la racine de l'arbre,
- le second niveau correspond à la liste des tonalités locales, c'est-à-dire des modulations, chacune d'entre elle étant reliée à la tonalité générale,
- le troisième niveau correspond à la séquence d'accords, telle qu'on pourrait la trouver sur une partition de jazz par exemple. Chacun de ces accords participe à une tonalité locale et lui est donc relié,
- le quatrième niveau correspond à la séquence de groupes de notes, qui constitue une réduction verticale de la polyphonie du morceau. Plusieurs groupes de notes pouvant sonner pendant le même accord, chacun d'entre eux est relié à un accord,
- le dernier niveau contient les notes de musique. Chacune d'entre elle participe à un groupe de notes, et lui est donc reliée.

La figure 40 présente un exemple d'arbre de l'harmonie proposé. La racine de l'arbre (tonalité principale) n'a pas été renseignée, car l'extrait choisi est très court. La tonalité principale est donc ici identique à l'unique tonalité locale.
On notera également que les notes doivent être considérées indépendamment les unes des autres (comme on le voit en bas à gauche). Afin de faciliter la mise en page de l'arbre, les notes ayant une même date de début et de fin ont été regroupées. On trouvera sur la figure 41 un second exemple d'arbre harmonie, où les groupes de notes sont cette fois annotés à la manière des accords. Les 5 niveaux d'information tonale sont ici présents.

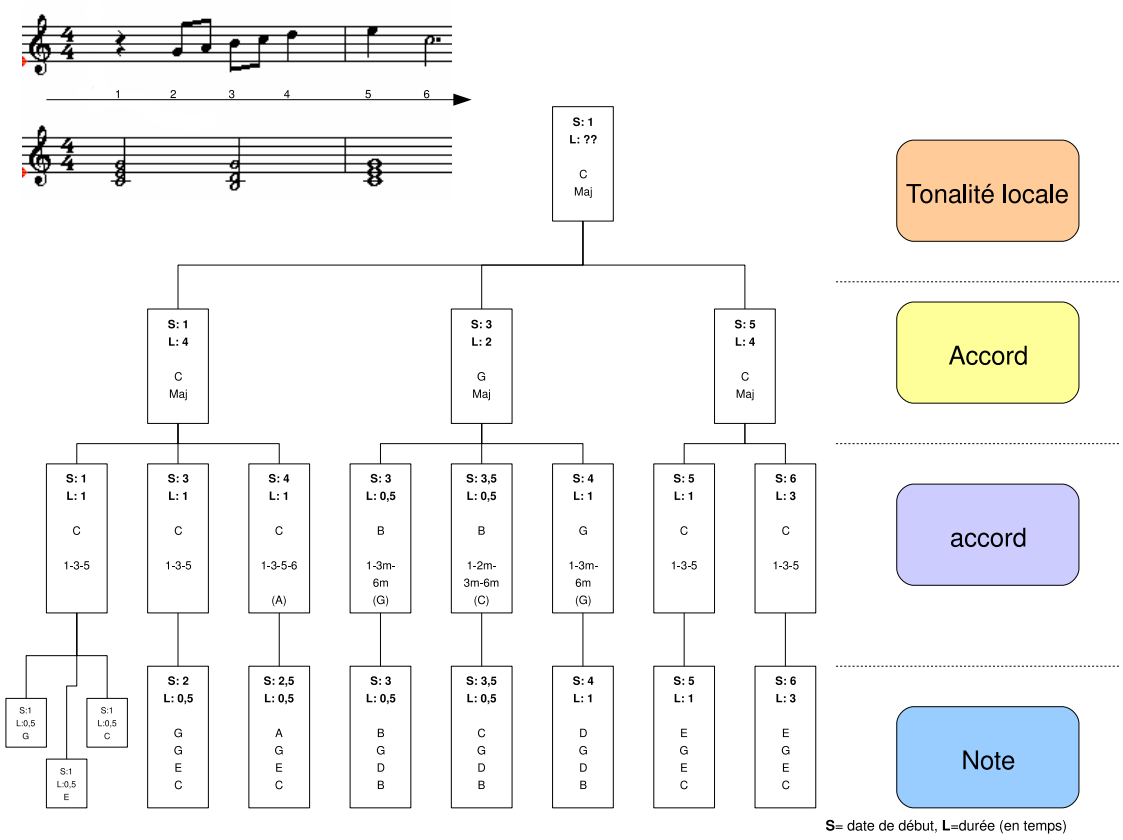


FIGURE 40 – Exemple d'arbre d'harmonie représentant un extrait musical.

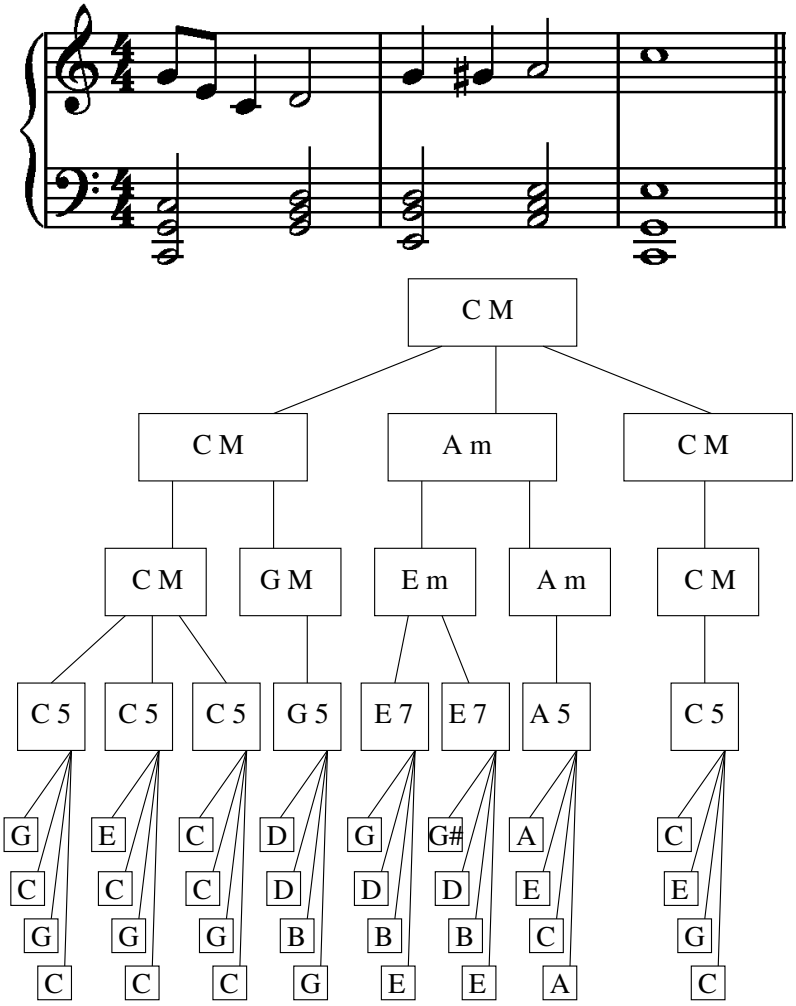


Figure 41 – Exemple de représentation différente d'un arbre d'harmonie complet d'un extrait musical.

3.3.2 Propriétés

Nous présentons ici plusieurs propriétés caractéristiques du modèle proposé. Contrairement aux autres représentations hiérarchiques de la littérature, cet arbre présente une profondeur fixe. Chaque niveau hiérarchique est ainsi associé à une information tonale unique :

- chaque sommet de cet arbre, à l'exception des feuilles, présente au moins un fils. Ainsi, un accord, par exemple, ne peut pas exister sans admettre au moins un groupe de notes, lui même composé d'au moins une note,
- chaque fils d'un sommet de niveau hiérarchique n admet un ou plusieurs fils de niveau hiérarchique $n+1$. Ainsi, une tonalité locale, par exemple, ne peut pas avoir de fils direct étant un groupes de notes, ou une note. Un fils d'une tonalité locale appartient donc au niveau hiérarchique accord.

Une autre propriété de cette arbre réside dans le fait qu'il est possible de lui associer un arbre relatif, où l'information n'est plus portée par les seuls sommets de l'arbre, mais également par les arêtes reliant les sommets. Ainsi, sur un tel arbre relatif, on pourrait représenter la date de début, la durée et le type d'un accord par un sommet, mais sa fondamentale serait représentée par un degré sur l'arête reliant cet accord à la tonalité locale associée.
L'arbre relatif associé à un arbre d'harmonie est invariant par transposition. Une transposition consiste à considérer un extrait d'une pièce musicale et à lui appliquer une transformation telle que les hauteurs h_i des notes de la pièce sont transformées en hauteurs h_j telles que l'intervalle (h_i, h_j) est constant. Les autres paramètres des notes ne sont pas modifiés.

Le chapitre 6 présente l'apport de ce nouveau modèle de l'harmonie dans le domaine de l'estimation de la similarité musicale.

Chapitre 4

Analyse des informations tonales

Dans les chapitres précédents, nous avons présenté et organisé cinq informations tonales distinctes : les notes de musique, les groupes de notes, les accords, les tonalités locale et les tonalités. Nous présentons ici différentes méthodes d'analyse de ces informations tonales : certaines de ces méthodes utilisent des propriétés musicales, d'autres des considérations statistiques ou des approches mettant en jeu une recherche de motifs. Pour chaque information, une distinction sera faite entre les méthodes d'analyse depuis une représentation symbolique de la musique, et les méthodes d'analyse depuis le signal audio. En plus de présenter les méthodes issues de la littérature, ce chapitre propose également des contributions pour l'analyse des accords et des tonalités depuis une représentation symbolique de la musique.

4.1 Notes et groupes de notes

Les notes de musique constituent un niveau de l'harmonie très bruité : toutes les notes jouées n'appartiennent pas à l'harmonie, et certaines peuvent ainsi constituer des ornements. Une estimation des accords et des tonalités peut ainsi être menée sans pour autant nécessiter d'estimation des notes de musique.

4.1.1 Analyse depuis une représentation symbolique

Dans la musique occidentale, une représentation symbolique propose généralement un codage des notes composant la pièce de musique. Qu'il s'agisse d'une partition musicale, ou d'un fichier MIDI (présenté en 2.2.1), les notes de musique peuvent ainsi être lues, et ne nécessitent pas d'estimation. Dans le cas des fichiers MIDI, il est possible d'utiliser des bibliothèques informatiques, comme la bibliothèque *midifile*[1], qui permettent d'accéder à

1. Disponible à l'adresse ftp ://ftp-archive.freebsd.org/pub/FreeBSD-Archive/old-releases/i386/1.0-RELEASE/ports/snd-util/fm/midifile/

la liste des événements MIDI, et donc des notes jouées pour chaque piste.

Pour analyser les groupes de notes, il est nécessaire de réorganiser les événements musicaux de la représentation symbolique, pour pouvoir représenter la pièce de musique sous la forme d'une séquence de groupes de notes. La première tâche consiste ainsi à déterminer les paramètres *date de début* et *durée* des groupes de notes. La transformation homorythmique (définie dans le chapitre 2 et illustrée sur la figure 7) peut en effet modifier la durée des notes, en transformant par exemple une noire en deux croches liées appartenant à deux groupes de notes différents. Il est important de souligner que tout changement dans l'harmonie donne lieu à un changement du groupe de notes : un groupe de notes peut ainsi débuter à une date où aucune note ne débute (voir figure 42).
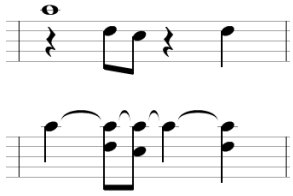


FIGURE 42 – Exemple de transformation homorythmique (extrait original en haut, transformation en bas). On observe que le troisième temps de la mesure correspond à un groupe de notes alors qu'aucune note ne débute à cette date.

La seconde tâche se situe dans la transformation des notes en classes de notes, c'est-à-dire en la réduction des notes présentes sur plusieurs octaves. Pour cela, les notes sont d'abord organisées du grave vers l'aigu, après quoi elles sont parcourues une à une. Pour chaque note rencontrée, deux cas sont à distinguer :

- aucune autre note de la même classe n'a encore été rencontrée, auquel cas sa classe est ajoutée aux classes du groupe de notes,
- au moins une autre note de la même classe a déjà été rencontrée, et la note est ignorée.

Enfin, la dernière tâche se situe dans l'estimation de la fondamentale. Nous proposons de considérer pour cela la minimisation de l'empilement de tierces nécessaire pour rencontrer toutes les classes de notes du groupe de notes. Nous avons ainsi choisi de considérer toutes les classes de notes du groupe de notes comme candidates au rôle de fondamentale. Nous construisons pour chacune d'entre elles les empilements de tierces nécessaires pour rencontrer toutes les autres classes de notes du groupe de notes considéré. On définit ensuite la hauteur de l'empilement comme le nombre de classes de notes visitées. La candidate pouvant présenter l'empilement de hauteur minimum est alors choisie comme la fondamentale. On notera que ces empilements de tierces sont constitués d'une alternance de tierces majeures (4 demi-tons) et de tierces mineures (3 demi-tons) : il faut en effet une tierce majeure et une tierce mineure (dans un ordre quelconque) pour former une quinte

(7 demi-tons) et ainsi pouvoir se déplacer sur le cercle des quintes pour atteindre toutes les classes de notes.

L'algorithme suivant présente la recherche de fondamentale parmi les classes d'un groupe de notes :

```
pour toute note x du groupe de notes
   initialiser l'empilement de tierces avec x comme fondamentale
   tierce_maj=vrai; a=0 ;
   tant que toutes les notes du groupe de notes n'ont pas été visitées
      si tierce_maj
         visiter la note située une tierce majeure vers l'aigu
      sinon
         visiter la note située une tierce mineure vers l'aigu
      tierce_maj=non tierce_maj; a=a+1;
   fin tant que
   noter a(x) la hauteur de l'empilement de tierces
   tierce_maj=faux; b=0;
   tant que toutes les notes de du groupe de notes n'ont pas été visitées
      si tierce_maj
         visiter la note située une tierce majeure vers l'aigu
      sinon
         visiter la note située une tierce mineure vers l'aigu
      tierce_maj=non tierce_maj; b=b+1;
   fin tant que
   noter b(x) la hauteur de l'empilement de tierces
   noter h(x)=min(a,b) la hauteur de l'empilement correspondant a x
fin pour
retourner la note x maximisant {h(y) \ y appartient au groupe de notes}
```

Considérons un exemple simple, avec le groupe de notes suivant : (Do, Fa, La) du grave vers l'aigu. Dans cet exemple, Do est donc la basse, et Fa la fondamentale. Les différents empilements de tierces donnent successivement :

- Do Mi Sol Si Ré Fa La : hauteur de l'empilement de tierces : 7,
- Fa La Do : hauteur de l'empilement de tierces : 3,
- La Do Mi Sol Si Ré Fa : hauteur de l'empilement de tierces : 7.

A l'aide de la minimisation de l'empilement de tierces, on trouve bien que Fa est la fondamentale.

4.1.2 Analyse depuis le signal audio

Depuis le signal audio, il est nécessaire de recourir à des méthodes de *transcription*, qui proposent une estimation des notes jouées dans la pièce analysée au fil du temps. Cette tâche de transcription est complexe. Elle débute par une description fréquentielle du signal

étudié, servant de support à une estimation des hauteurs présentes dans l'extrait analysé. Nous présentons dans un premier temps les méthodes de transcription monophonique (où une seule note est jouée à un instant donné), avant de nous pencher dans un second temps sur les méthodes de transcription polyphonique, où plusieurs notes peuvent être jouées en même temps.

4.1.2.1 Transcription monophonique

Les premières méthodes de transcription monophonique datent de la fin des années 1960 [Nol67, RSC+74], et visent à détecter les hauteurs perçues dans l'extrait analysé. Ces méthodes, reprises par des travaux plus récents [BMS00, MS00], utilisent en particulier l'auto-corrélation du signal étudié. La fonction d'"auto-corrélation A du signal numérique x supposé centré en 0, et sur un intervalle de temps correspondant à N échantillons est définie par :

$$A(i) = \sum_{k=0}^{N-i} x_k x_{k+i}$$

où N représente la taille de la fenêtre sur laquelle cette auto-corrélation est calculée. En recherchant le maximum sur i de cette auto-corrélation (en excluant $i = 0$), on trouve une valeur i_{max}, permettant de trouver la fréquence correspondant à ce maximum et donc de donner la fréquence f estimée en fonction de la fréquence d'échantillonnage f_{ech} du signal : $f = f_{ech}/i_{max}$
Ces méthodes de transcription peuvent présenter plusieurs limites : le confusion entre f et $2f$ ou $0.5f$, la faible robustesse au bruit, et surtout, l'absence de prise en compte de la polyphonie, c'est-à-dire du fait que plusieurs notes peuvent être jouées en même temps.

4.1.2.2 Transcription polyphonique

Les méthodes de transcription polyphonique proposent une estimation des différentes hauteurs perçues simultanément sur une période de temps donnée. On distingue dans la littérature deux types d'estimation de hauteurs : les estimations itératives et les estimations jointes (détaillées ci-dessous). La transcription polyphonique représente un domaine de recherche à part entière, dont l'étendue dépasse le cadre des travaux présentés dans ce manuscrit. Pour de plus amples informations et une présentation plus détaillée des méthodes de transcription polyphonique, on pourra notamment se référer à l'ouvrage de Klapuri et Davy [KD06].

Estimations itératives

Ces méthodes reposent sur une estimation successive des différentes hauteurs. Chaque hauteur est dans un premier temps détectée et ajoutée à la liste des hauteurs estimées.

Dans un second temps, il est nécessaire de supprimer les fréquences correspondantes (fondamentale et harmoniques) du signal restant à analyser. Ces deux étapes sont à répéter tant qu'un critère d'arrêt n'est pas atteint (comme le nombre de hauteurs à détecter, par exemple).
Parmi ces méthodes, on trouve les travaux de Karjalainen et Tolonen [KT99], de Ortiz-Berenguer [OBCqTG05] et Klapuri [Kla03, Kla08]. Elles diffèrent dans la manière d'estimer une note à partir du signal analysé, le critère d'arrêt choisi et surtout l'estimation des contributions fréquentielles d'une hauteur détectée, pour pouvoir soustraire les fréquences correspondantes au signal restant à analyser. Ce dernier point constitue l'enjeu majeur des méthodes d'estimation itérative. En 2001, Klapuri propose une méthode reposant sur la notion d'enveloppe spectrale de la hauteur détectée, illustrée sur la figure 43. Les pics de la courbe en trait épais correspondent aux fréquences harmoniques d'une hauteur venant d'être détectée, alors que les pics en trait fin représentent les fréquences du signal restant à analyser. Klapuri remarque qu'il est possible de relier les pics en trait épais par une enveloppe (en trait fin) gardant un profil constant, quels que soit les hauteurs et les instruments considérés. Cette enveloppe spectrale permet ainsi d'estimer les fréquences à soustraire au signal considéré lorsqu'une hauteur vient d'être détectée. Les pics en trait fin situés au-dessus de cette enveloppe spectrale correspondent ainsi aux fréquences qui seront conservés après soustraction, et qui permettront de détecter une autre hauteur. Cette méthode, baptisée *spectral smoothness* fait référence depuis lors dans ce domaine [Kla01]. Néanmoins, en plus de la difficulté d'établir un critère d'arrêt, cette méthode, comme toutes les méthodes itératives, utilise une approximation de l'amplitude des harmoniques correspondant aux hauteurs détectées, ce qui représente donc une source potentielle d'erreur.

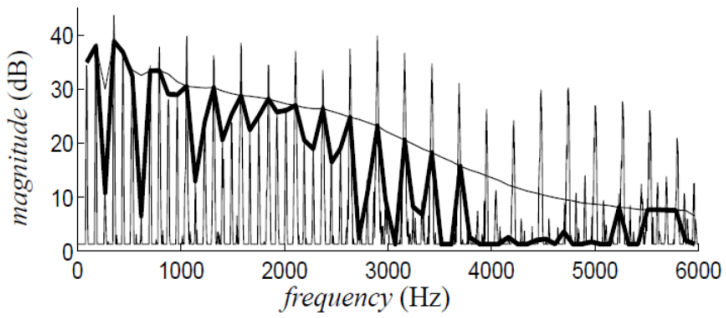


Figure 43 – Illustration du *spectral smoothness*, d'après [Kla01]. L'enveloppe est en trait fin, et relie les pics de fréquences (en trait épais) correspondant à une hauteur venant d'être détectée. Les pics de trait fin situés au-dessus de cette enveloppe seront conservés après soustraction des fréquences correspondant au trait épais.

Estimations jointes

Les méthodes d'estimation jointe se proposent d'estimer simultanément l'ensemble des notes jouées durant l'extrait considéré. Le principe est ici d'utiliser des filtres en peigne pour supprimer les pics spectraux correspondant aux fréquences harmoniques des hauteurs estimées (trait épais sur la figure 43). Un critère énergétique (comme l'énergie spectrale) est ensuite calculé, et doit atteindre un minimum lorsque les fréquences choisies sont correctes, et que celles-ci correspondent donc bien aux hauteurs présentes dans l'extrait analysé. Tel est le principe des méthodes proposées par Tolonen [TK00] et de Cheveigné [dCB03] au début des années 2000.
Ces méthodes résolvent le problème de la soustraction approximative des contributions des hauteurs successivement analysées, rencontré dans les méthodes d'estimations itératives. Cependant, les estimations jointes imposent de fixer un degré de polyphonie maximal (nombre de notes jouées en même temps), et ne sont efficaces que pour les polyphonies restreintes, présentant un faible nombre de hauteurs jouées simultanément (2 ou 3).

Que ce soit avec des méthodes itératives ou jointes, la transcription polyphonique ne peut être réalisée sans une certaine part d'erreur, et leurs performances dépendent fortement du nombre des sources musicales et de la nature du timbre supposés [EDB07].

4.2 Accords

Cette section présente des méthodes d'estimation d'accords, qui constituent une élément central de la description de l'harmonie par la communauté MIR *Music Infomation Retrieval*. La séquence d'accords d'un morceau fournit en effet une représentation particulièrement intéressante, notamment pour les styles jazz et pop. L'analyse de l'harmonie d'un morceau de musique par sa séquence d'accords constitue également une problématique de plus en plus importante pour les théoriciens de la musique [BWF11]. Dans un premier temps, nous nous penchons sur l'analyse depuis une représentation symbolique de la musique, en présentant des méthodes existantes avant de proposer certaines améliorations. Nous étudions ensuite l'analyse des accords depuis le signal audio.

4.2.1 Depuis une représentation symbolique de la musique

Nous nous restreignons dans cette section aux méthodes d'analyse des accords depuis une représentation symbolique de la musique. Après avoir présenté les méthodes de la littérature, nous proposons une méthode directe de reconnaissance d'accords, reposant sur une comparaison de profils de notes avec des profils d'accords. Nous présentons ensuite des améliorations pur cette méthode, en tenant compte des propriétés de l'harmonie de la musique.

4.2.1.1 Méthodes existantes

Maxwell propose en 1992 [Max92] une méthode d'estimation d'accords utilisant plus de 50 règles. La pièce de musique est d'abord représentée sous la forme d'une séquence de verticalités (*verticals*), proche de notre définition des groupes de notes. Les règles visent ensuite à déterminer quelles verticalités sont des accords, et quelles autres sont des ornements. Une de ces règles indique par exemple qu'une verticalité composée d'une unique note constitue un ornement du dernier accord rencontré (Temperley affirmera plus tard que ce choix est discutable [Tem99a]). La plupart des règles reposent sur les notions de consonance et dissonance, et deviennent difficilement applicables lorsque ces dernières évoluent au fil des styles musicaux. Ainsi, une consonance en musique classique peut être différemment définie qu'une consonance en jazz.
Temperley, en 1999 [Tem99a], propose une méthode reposant également sur des règles, mais inspirées cette fois de la théorie génératrice de la musique tonale de Lerdahl et Jackendoff [LJ85]. Cette méthode réalise tout d'abord une analyse de la métrique rythmique de la pièce étudiée, en distinguant les temps forts des temps faibles. Une des règles permettant cette analyse consiste par exemple à privilégier la création de temps forts aux dates de début des notes de durées importantes. Une fois l'analyse de la métrique réalisée, la méthode définit une segmentation de la pièce étudiée, chacun de ces segments faisant ensuite l'objet d'une estimation d'accord. Pour définir la segmentation, les règles proposées par Temperley privilégient par exemple la création d'un nouveau segment sur un temps fort. Enfin, l'estimation d'accord sur un segment donné a lieu, et propose pour la fondamentale de l'accord estimé. Les règles permettant cette estimation privilégient par exemple une proximité sur le cycle des quintes entre fondamentales successives. Temperley propose une évaluation de cette méthode sur plusieurs pièces classiques. Il juge les résultats "raisonnables" et met en évidence quelques points défaillants de cette méthode, en s'appuyant notamment sur l'analyse de la Gavotte de la suite française numéro 5 de J.S. Bach : selon lui, cette méthode n'identifie pas correctement certaines notes ornementales, leur conférant ainsi un rôle tonal et provoquant certaines erreurs lors de l'application des règles.
Plus récemment, Pardo [PB02] propose un autre algorithme, reposant lui sur la définition de segments musicaux, et sur le calcul de scores pour chacune des triades majeures et mineures sur chacun de ces segments. Bien que les termes *corrélation* ou *profils de notes* ne sont jamais employés, une partie de la méthode proposée par Pardo est donc directement inspirée de l'algorithme KS, que l'on décrira section 4.3.1, et qui compare un profil de notes à des profils de référence. Les profils de notes sont également utilisés dans les travaux de Rhodes [RLM07] et dans ceux de Paiement [PEB05], qui utilisent des modèles statistiques (les modèles statistiques formant la majorité des travaux existants en estimation d'accords depuis le signal audio, ils seront décrits en section 4.2.2).

4.2.1.2 Méthode directe : *MD*

Une première méthode directe peut être proposée, en utilisant la description de l'harmonie offerte par les profils de notes. Cette méthode directe comprend trois étapes :

(1) *La segmentation de la pièce analysée.* Nous choisissons d'adopter dans cette méthode une segmentation uniforme et de résolution caractéristique de la durée d'un accord [RLM07]. La mesure est choisie comme unité de segmentation.

(2) *La constitution du profil de notes.* Les profils de notes sont des vecteurs de dimension 12, mesurant l'influence des 12 classes de notes jouées dans le segment étudié. Ce profil est construit en sommant les durées des notes du segment selon leurs classes, et en sommant leurs durées, exprimées en fonction d'une unité de temps. Si nous considérons un segment dans lequel trois notes sont présentes : un Do (blanche), un Sol (noire) et un Si (croche), et que nous choisissons comme unité de temps élémentaire la croche, le profil de notes construit sur ce segment sera :

$$p = (4, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 1)$$

On trouvera une illustration de construction d'un profil de notes sur la figure 44.

(3) *L'estimation de l'accord.* Nous choisissons de considérer 24 accords (12 de type majeur et 12 de type mineur), et de construire pour chaque accord des profils en triade. Ces profils sont composés de la fondamentale de l'accord, ainsi que des notes à la tierce (majeure ou mineure selon le mode) et à la quinte (voir figure 45). Le profil de notes du segment est ensuite comparé aux profils d'accords. Ces profils sont comparables à des profils de notes, et sont construits à l'aide de considérations musicales. On compare le profil de notes obtenu sur un segment aux profils d'accords en calculant une valeur de corrélation entre les 24 accords de triades et le segment étudié. L'accord dont le profil présente la plus forte valeur de corrélation constitue l'accord estimé sur le segment considéré. Dans la suite, le calcul de corrélation est réalisé en effectuant un produit scalaire entre les deux profils comparés.

Cette méthode directe est par la suite nommée *MD*. Elle est évaluée et comparée à une autre méthode dans la section suivante.

4.2.1.3 Méthode utilisant les groupes de notes : *MGN*

Dans la méthode directe décrite précédemment, le profil de chaque segment est construit à partir des notes présentes dans le segment, sans faire de distinction quant à leurs importances respectives dans l'harmonie. D'un point de vue musical, il semble pourtant

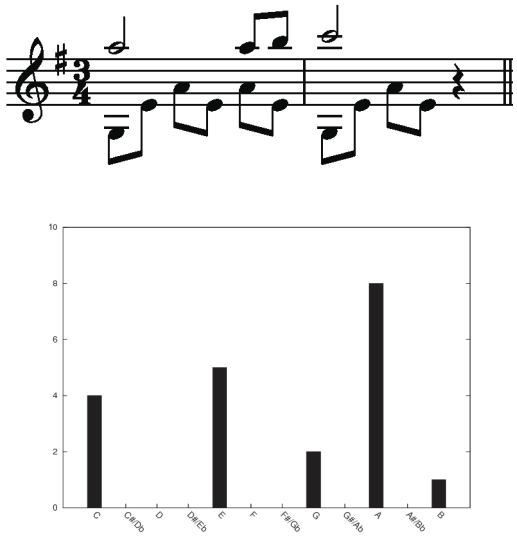


FIGURE 44 – Illustration de la construction d'un profil de notes. Sur un segment comportant les notes de la figure du haut, le profil obtenu est représenté sur la figure du bas (avec la pondération des 12 classes de notes en ordonnée). On notera que le poids de chacune des notes dépend de sa durée. On définit ainsi un poids de 1 pour une croche, 2 pour une noire, 4 pour une blanche, etc.

intéressant de conférer un plus grand rôle aux notes sonnant simultanément, car elles constituent l'aspect vertical de la musique, c'est-à-dire l'harmonie. Il a ainsi été choisi d'utiliser une approche qui tenait davantage compte de l'harmonie, en considérant les propriétés des groupes de notes de chaque segment pour construire le profil de notes de ce segment. Ces considérations auront pour effet de modifier les contributions des notes présentes dans le segment, qui ne seront donc plus seulement fonction de leurs durées. Voici les deux modifications que nous proposons, par rapport à la méthode *MD* :

- plus un groupe de notes est riche (plus il comporte de notes), plus les notes qui le composent contribuent à l'harmonie. La contribution de chaque note doit donc dépendre du nombre de notes du groupe de notes auquel elle appartient,
- pour chaque groupe de notes, la fondamentale joue un rôle prépondérant sur l'harmonie par rapport aux autres notes. La contribution de cette dernière doit être plus importante.

Nous proposons donc une nouvelle méthode, nommée *MGN* (Méthode Groupes de Notes). Si $c_{\mathrm{MD}}(i)$ représente la contribution d'une note i avec la méthode *MD*, on définit $c_{\mathrm{MGN}}(i)$ comme suit : $c_{\mathrm{MGN}}(i) = c_{\mathrm{MD}}(i)n(i)f(i)$ avec $n(i)$ le nombre de notes présents de le même groupe de notes que i, et $f(i) = 2$ si i est fondamentale de son groupe de notes, $f(i) = 1$ sinon.

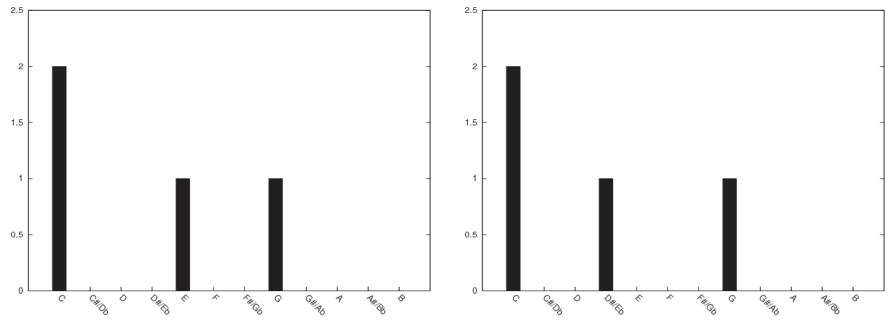


FIGURE 45 – Profils d'accords en triade de Do Majeur (à gauche) et Do mineur (à droite). Les profils d'accords des autres accords en triade s'obtiennent par rotation d'indice. Nous avons choisi de donner un poids double à la fondamentale.

Contrairement à $c_{\mathrm{MD}}(i)$, qui ne dépend que de la durée de la note i, $c_{\mathrm{MGN}}(i)$ tient donc en plus compte des propriétés du groupe de notes de i.

Comme dans la méthode *MD*, pour chaque profil de notes, la méthode *MGN* cherche le maximum de corrélation avec les mêmes profils d'accords (en triade).

Exemple

Considérons l'extrait musical représenté figure 46, qui représente une mesure en Do Majeur suivie de deux profils de notes pour cette mesure :

- le premier, à gauche, obtenu avec la méthode *MD* : chaque note du segment est ajoutée au profil de notes, avec un poids dépendant de sa durée (1 pour une croche, 2 pour une noire, 4 pour une blanche, etc),
- le second, à droite, obtenu avec la méthode *MGN* : chaque note du segment est ajoutée au profil de notes, avec un poids dépendant non seulement de sa durée, mais également du nombre de notes jouées en même temps que cette note, et du fait qu'elle soit ou non la fondamentale d'un groupe de notes.

Ainsi, avec la méthode *MGN*, le Do grave figurant dans l'accord de blanches au début de la mesure n'a plus un poids de 4 (correspondant à la durée d'une blanche), comme dans la méthode *MD*, mais un poids de $4 \times 3 \times 2 = 24$, car le groupe de notes comporte trois notes, et que ce Do est la fondamentale du groupe de notes. Le Fa entre le troisième et le quatrième temps conserve un poids inchangé entre les deux méthodes (1 dans les deux cas). On constate donc que la méthode *MGN* a tendance, sur cet exemple, à renforcer considérablement l'importance harmonique du Do dans cette mesure. Sur le profil correspondant à la méthode *MGN*, on note ainsi que le Do a un poids plus élevé que le Sol, alors qu'il apparaît moins que ce dernier dans la mesure (2 temps pour le Do, contre 2 temps et 1 demi-temps pour le Sol). On remarque que ce deuxième profil présente une forme bien plus

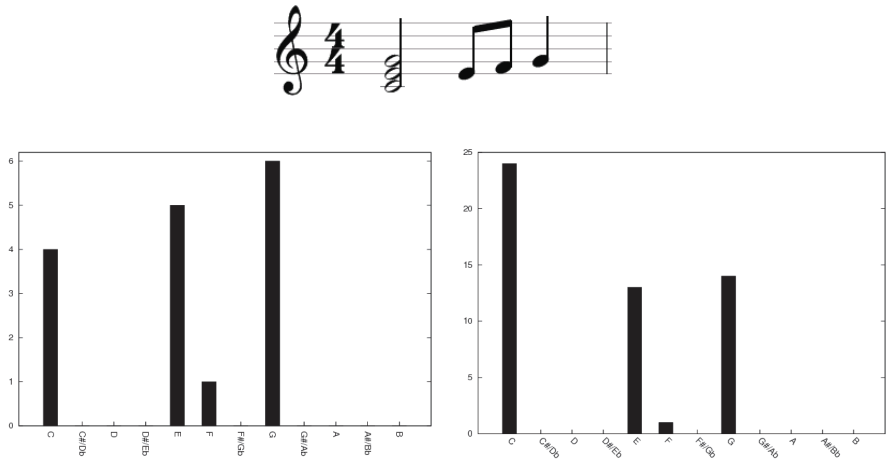


Figure 46 – Une mesure en Do Majeur, suivie de la comparaison de profils de notes de cette mesure selon leurs pondérations. À gauche se trouve le profil de notes d'après la méthode *MD*. À droite se trouve le profil de notes obtenu avec la méthode *MGN*. On remarque que ce deuxième profil présente une forme plus proche du profil en triade de Do Majeur que le profil obtenu avec la méthode *MD*.

proche du profil en triade de Do Majeur que le profil obtenu avec la méthode *MD*, ce qui donne un meilleur score de corrélation avec le profil de référence Do Majeur pour le profil issu de la méthode *MGN* que pour le profil de notes issu de la méthode *MD*.

La figure 47 présente les résultats du calcul des corrélations entre les profils de notes illustrés figure 46, et les 24 profils d'accords en triade. Les deux profils de notes ont été normalisés pour ce calcul, afin que chacun présente une somme de ses composantes égale à 1.

On remarque ainsi que le score de corrélation correspondant au Do Majeur est meilleur pour le profil de notes issu de la méthode *MGN* (1,44 contre 1,19 pour le profil de notes issu de la méthode *MD*). De plus, la méthode *MGN* permet d'augmenter l'écart entre le score de Do Majeur et le deuxième meilleur score : la différence avec le deuxième meilleur score (Do mineur) est de 0.25 pour la méthode *MGN*, alors qu'elle est de 0.19 entre Do Majeur et Mi mineur pour la méthode *MD*. Enfin, on remarque que la méthode proposée permet de faire baisser les corrélations relatives aux accords éloignés de l'accord à déterminer : 6 accords présentent des scores au-dessus de 0.6 (dont 3 au-dessus de 0.8) avec la méthode *MD*, alors que seulement 4 sont au-dessus de 0.6 (dont 2 au-dessus de 0.8) avec la méthode *MGN*.

Si l'étude de l'exemple précédent illustre l'intérêt de la prise en compte des informations tonales issues des groupes de notes pour déterminer les accords, une évaluation de la détection d'accords doit être envisagée sur une plus grande quantité d'accords, et sur des morceaux de musique entiers. Cette évaluation est présentée ci-après.

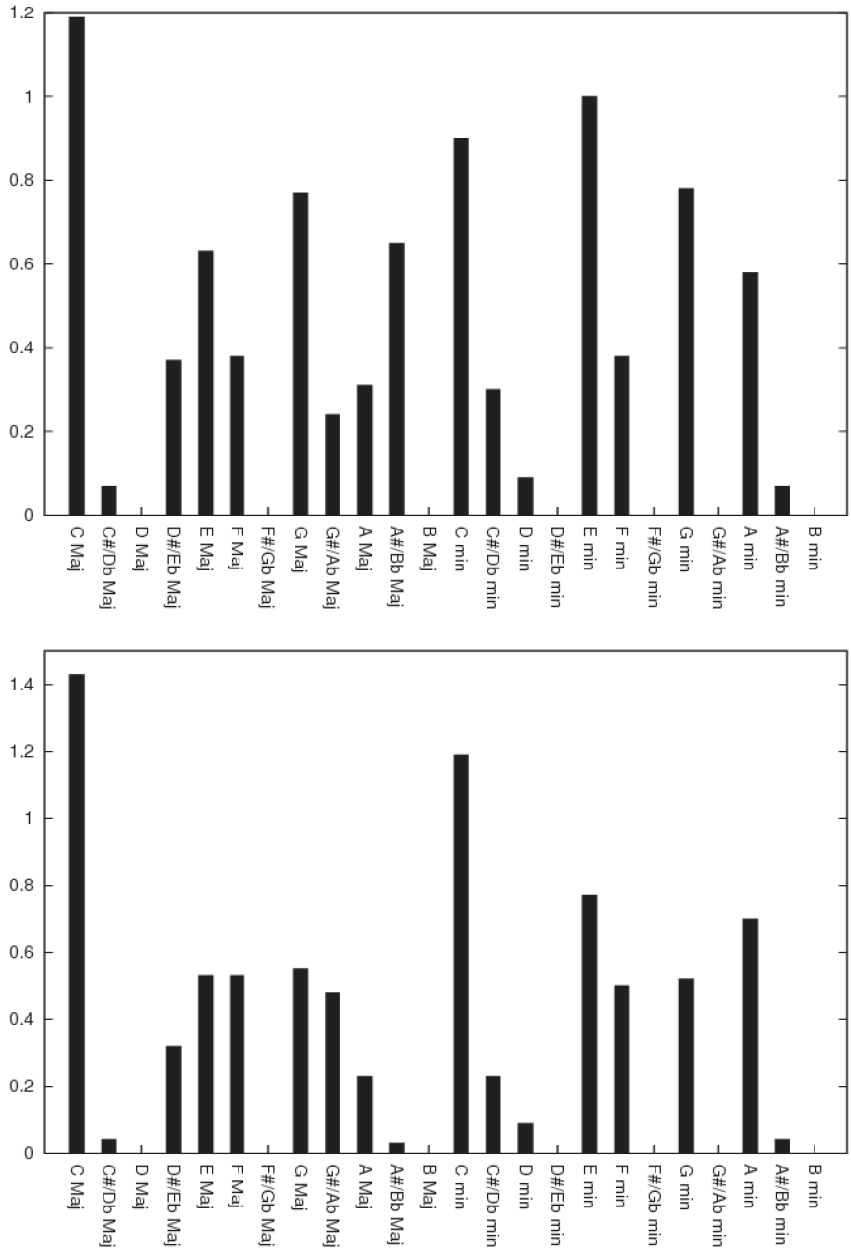


FIGURE 47 – Résultats des calculs de corrélations. En haut se trouvent les scores de corrélation du profil de notes de la méthode MD avec les 12 accords majeurs et les 12 accords mineurs. En bas sont représentées les mêmes scores pour le profil de notes issu de la méthode MGN. On remarque que ce dernier renforce la corrélation correspondant à l'accord à estimer (ici, Do Majeur).

Artiste	Animals	Radiohead	Otis Redding	The Eagles
Titre	House Of The Rising Sun	Creep	Dock Of The Bay	Hotel California
MD	94.19	93.33	80.88	93.27
MGN	**98.26**	**95.56**	**100**	**99.33**
Artiste	Pink Floyd	Pink Floyd	Pink Floyd	The Beatles
Titre	The Wall	Wish You Were Here	Comfortably Numb	Can't Buy Me Love
MD	88.61	**91.14**	80	**100**
MGN	**89.87**	89.87	**98.5**	98.75
Artiste	Barbara	The Beatles	The Beatles	The Beatles
Titre	L'Aigle Noir	Magical Mystery Tour	Come Together	Ticket To Ride
MD	92	94.74	86.47	**97.83**
MGN	**100**	94.74	**91.26**	96.52

Méthode	Moyenne pondérée par titre	Moyenne pondérée par mesure
MD	91.04	90.28
MGN	**96.06**	**96.44**

Table 10 – Résultats de précision des méthodes d'estimation d'accords MD et MGN (en % d'accords corrects estimés).

Bases de données et protocole d'évaluation

Nous avons sélectionné 12 morceaux pop (voir annexe A.1.1), et annoté manuellement les fichiers MIDI correspondant à l'aide d'accords majeurs et mineurs. Ces morceaux présentent tous un changement d'accord (au plus) par mesure. Chaque mesure de chaque morceau donne ainsi lieu à une évaluation des accords estimés, pour un total de 1195 mesures sur la base de données.

Pour chaque mesure, nous comparons l'accord annoté avec l'accord estimé, et ce pour chacune des méthodes. Si l'accord annoté est identique à l'accord estimé (même fondamentale et même mode), le score sur la mesure est de 1. Sinon, le score de la mesure est 0. Le score de chaque morceau est ensuite obtenu en divisant le nombre de mesures donnant lieu à une estimation correcte par le nombre total de mesures du morceau. Nous proposons également des scores moyens sur la base de données, en procédant soit au calcul de la moyenne des scores pour chaque morceau (moyenne pondérée par titre), soit en pondérant le score de chaque morceau par son nombre de mesures (moyenne pondérée par mesure).

4.2.1.4 Résultats

On trouvera les résultats de l'estimation des accords dans le tableau 10. On remarque que la méthode *MGN* améliore significativement les résultats, en réalisant une hausse d'au moins 5% des résultats, que ce soit en considérant la moyenne par morceau ou par accord. Les résultats de la méthode *MD* réalisant déjà une score élevé (plus de 90%), il est à noter

qu'un écart moyen de 5% apparaît comme important. On notera également que, sur une minorité de morceaux, la méthode *MGN* est moins précise que la méthode *MD* (comme pour *Can't Buy Me Love* ou *Ticket To Ride*), la différence n'excédant toutefois jamais 2%). En revanche, sur les titres où la méthode *MGN* est plus précise que la méthode *MD*, les résultats sont améliorés de manière plus significative : 18,5% d'amélioration sur *Comfortably Numb*, et presque 20% pour *Dock Of The Bay*, par exemple.

Il est possible d'avancer une explication quant à la possible introduction d'erreurs de la méthode *MGN*, en s'appuyant sur le fait que la différence principale entre les deux méthodes réside dans la suppression des notes appartenant à la même classe, et au fait qu'un poids plus important est conféré à la fondamentale dans chaque groupe de notes. Considérons le segment constitué des notes suivantes (du grave vers l'aigu) :
(Do Mi Sol Do Do Do Do Do)
Les deux méthodes estiment correctement l'accord de Do Majeur, car le groupe de notes associé est (Do Mi Sol), avec une basse de do et une fondamentale de do. En revanche, si l'on ajoute un instrument jouant simplement un La, alors le groupe de notes devient (Do Mi Sol La) et la fondamentale devient La. La méthode *MGN* indique alors un accord de La mineur, tandis que la méthode *MD* continue d'estimer un accord de Do Majeur (car la seule note La ne présente pas de grande influence sur la répartition statistique des notes de l'accord, étant donné que les nombreux o ne sont pas regroupés en une unique classe de notes). Lorsque l'accord correct est Do Majeur, il peut ainsi se retrouver estimé en La mineur par l'amélioration proposée dans la méthode *MGN*. Il serait donc possible d'envisager une future amélioration de cette méthode, en testant par exemple la corrélation avec Do Majeur si un accord de La mineur est détecté. Si la corrélation avec Do Majeur est relativement proche de avec La mineur, on pourrait alors recourir à un autre type d'information, comme la considération de la tonalité estimée, pour choisir entre ces deux accords.

4.2.1.5 Conclusion

La prise en compte des groupes de notes, et de l'information tonale ainsi apportée, semble être bénéfique à l'estimation des accords. Cette constatation constitue un premier pas vers l'hypothèse que des informations tonales de plus haut niveau que les notes de musique proposent une description de l'harmonie moins bruitée, et permettent une estimation plus précise des autres informations tonales. Nous proposons dans la suite d'autres expériences mettant en avant cette hypothèse.

4.2.2 Depuis le signal audio

Après nous être attardés sur l'estimation d'accords depuis une représentation symbolique de la musique, nous nous penchons ici sur les méthodes d'estimation d'accords depuis le signal audio. Comme nous l'avons vu précédemment, ces méthodes sont significativement différentes, car il semble difficile d'utiliser une étape de transcription polyphonique des enregistrements étudiés, en raison des erreurs induites par les méthodes de transcription. Les méthodes d'estimation d'accords depuis le signal audio utilisent ainsi la description de l'harmonie apportée par les vecteurs chroma, que nous avons présentés en 2.2.2.

Parmi ces méthodes, trois grandes catégories peuvent être distinguées dans la littérature :

- les méthodes utilisant une recherche et une reconnaissance de modèles d'accords,
- les méthodes fondées sur un apprentissage statistique, employant le plus souvent un modèle de Markov caché,
- les méthodes hybrides, où la théorie musicale est utilisée comme complément à l'approche statistique.

4.2.2.1 Recherche de modèles d'accords

Nous avons vu précédemment que certaines méthodes d'estimation d'accords depuis une représentation symbolique de la musique utilisent une comparaison entre les profils de notes observés et des profils d'accords. De la même manière, les méthodes de recherche de modèles d'accords comparent les vecteurs chroma avec des vecteurs caractéristiques des accords recherchés, appelés vecteurs modèles d'accords. Le vecteur modèle le plus "proche" du chroma étudié donne alors l'accord estimé sur la période de temps considérée.
Fujishima, en 1999 [Fuj99], fut le premier à proposer une telle méthode, en utilisant des vecteur modèles binaires (composés uniquement de 1 et de 0), et en distinguant 27 types d'accords différents. Les vecteurs chromas utilisés sont de dimension 12. L'efficacité de cette méthode s'avère satisfaisante en estimant des accords joués depuis un synthétiseur MIDI, avec une précision supérieure à 90% d'accords correctement estimés. Néanmoins, évaluée sur de la musique réelle (en opposition à la musique générée par un synthétiseur MIDI), la méthode commet de nombreuses erreurs.

Ces erreurs sont dues au problème d'accordage, qui constitue une contrainte importante dans le traitement du signal [KO09] : sur un enregistrement réel, la fréquence de référence du La peut ne pas être exactement égale à 440Hz. Meinard Müller a ainsi communiqué à la communauté [2] le fait que certains enregistrements des Beatles présentaient une fréquence de référence différente de 440Hz. Ces différences d'accordage sont appelées déviations, et sont mesurées en demi-tons. On trouvera la répartition de ces déviations sur le tableau 11.

2. via la liste de diffusion MUSIC-IR music-ir@listes.ircam.fr

Déviation (en demi-tons)	0.0	0.25	0.33	0.50	0.67
Nombre de morceaux	77	30	58	7	2

TABLE 11 – Déviations de l'accordage sur les enregistrements des Beatles, d'après Meinard Müller.

En 2005, Harte et Sandler [HS05] proposent une méthode similaire à celle de Fujishima, mais considèrent quatre types d'accords au lieu de deux (majeurs, mineurs, diminués et augmentés) et intègrent un algorithme pour gérer les problèmes d'accordage à l'aide de vecteurs chroma de dimension 36. Chaque classe de notes correspond ainsi à 3 composantes différentes (au lieu d'une composante vectorielle par classe de notes pour les vecteurs chroma de dimension 12). En observant les maxima locaux de tous les vecteurs chroma de la pièce étudiée, ils déterminent où se situent les frontières entre classes de notes dans les vecteurs chroma de dimension 36, et les transforment ensuite en vecteurs de dimension 12.

Lee [Lee06], en 2006, propose une méthode légèrement différente dans le calcul des vecteurs chroma. Il considère ainsi non plus directement le résultat de la transformée de Fourier, mais le produit harmonique spectral (*Harmonic Product Spectrum*), comme proposé par Scholes [Sch68]. Deux types d'accords différents (majeur et mineur) sont considérés. Pour chaque trame de signal, la maximisation de la corrélation avec des modèles d'accords binaires donne l'accord estimé.

En 2009, Oudre [OGF09] propose une étude sur l'influence des différents profils de référence pour la reconnaissance d'accords depuis le signal audio. Il remarque également que les méthodes décrites ci-dessus utilisent toutes un lissage de la séquence d'accords estimée, corrigeant par exemple une sous-séquence d'accords de la forme *aaabaaa* en séquence *aaaaaaa*. Il propose d'effectuer ce lissage non pas sur la séquence d'accords estimées, mais sur le chromagramme, en réalisant une médiane pour chaque composante des vecteurs chroma. Il met ainsi en évidence l'apport du filtre médian sur le traitement du chromagramme.

4.2.2.2 Modèles statistiques

Les méthodes utilisant une approche statistique constituent une grande part des travaux existants dans le domaine de la reconnaissance d'accords depuis le signal audio. Elles ont la plupart du temps recours à un modèle de Markov caché (ou HMM) [Rab90], illustré figure 48. Si une séquence d'accords est considérée comme une chaîne de Markov, la probabilité d'occurrence de l'accord A pour une trame i dépend uniquement de l'accord estimé sur la trame précédente, et est donc exprimée sous la forme $P(A_i|A_{i-1})$. Ces modèles ont

ainsi recours à une matrice de transition M, ayant pour dimension le nombre fini d'accords considérés. Dans le cas où l'on cherche à estimer des accords de triades, et où 24 accords sont donc considérés, les matrices de transitions ont une taille 24x24, et sont telles que $P(A_i|A_{i-1}) = M_{j,k}$, où j est l'indice correspondant à l'accord A_i, et k l'indice correspondant à A_{i-1}. Dans un HMM, les états ne sont pas directement accessibles, et doivent être décrits à l'aide d'observations, qui sont réalisées sur chaque trame en calculant un vecteur chroma. Un vecteur chroma X est modélisé comme étant une variable aléatoire dépendant uniquement de l'état (et donc de l'accord A) sur la trame considérée, et ce, suivant la distribution $P(X_i|A_i)$.

La modélisation selon un HMM implique ainsi :
- une description de l'accord A_i à estimer sur une trame i en fonction du vecteur chroma X_i, modélisé par $P(X_i|A_i)$,
- une description de la probabilité d'occurrence de A_i étant donné l'accord estimé sur la trame précédente, modélisée par $P(A_i|A_{i-1})$.

La séquence d'accords $a^*_{1:N}$ estimée entre les trames 1 et N est finalement exprimée à partir du chromagramme $x_{1:N}$ comme :

$$a^*_{1:N} = \arg\max_{a_{1:N}} P(a_{1:N}|x_{1:N})$$

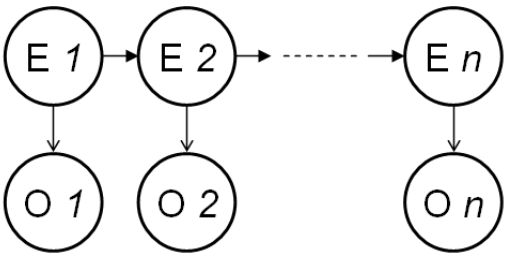


Figure 48 – Illustration d'un Modèle de Markov Caché (HMM). Les états (E) sont les accords à estimer, et sont décrits à l'aide d'observations (O) : les vecteurs chroma.

Ces méthodes nécessitent la plupart du temps une phase d'apprentissage pour tirer un meilleur parti de l'approche statistique. Avant de pouvoir estimer la séquence d'accords d'un morceau inconnu, un entraînement du modèle doit ainsi être mené sur une base d'apprentissage. Un certain nombre de morceaux doivent donc être annotés, après quoi la phase d'apprentissage permet de définir les probabilités de transitions entre accords, et les probabilités d'occurrences des accords en fonction des profils des vecteurs chroma.

Le premier modèle de ce type appliqué à la reconnaissance d'accords fut proposé par Sheh et Ellis en 2003 [SE03]. Dans ce modèle, une phase d'apprentissage utilisant espérance-maximisation [DLR77] est tout d'abord utilisée, pour définir la moyenne et la variance des

vecteurs modèles d'accords, à partir d'enregistrements audio annotés. Cette phase d'apprentissage permet également de définir la matrice de transition entre accords. L'algorithme de Viterbi [Vit67] est ensuite utilisé pour obtenir la séquence d'accords à partir du chromagramme. Les auteurs montrent que l'adaptation des HMM au domaine de la reconnaissance d'accords est possible. Ils affirment cependant que lorsqu'un enregistrement musical inconnu est considéré, la précision de la méthode n'est pas satisfaisante.

Les travaux de Ryynänen et Klapuri [RK08] reprennent un principe similaire, mais utilisent un profil d'accord grave (correspondant aux fréquences et aux notes graves) combiné à un profil d'accord aigu (pour les fréquences, et donc les notes aiguës). Ces deux profils sont utilisés pour tenter de distinguer la fondamentale des accords, en utilisant le profil grave, du type de l'accord (donné par la tierce et la quinte, par exemple), en utilisant le profil aigu. La séparation entre les registres grave et aigu se situe à environ 140Hz et correspond à la note de Ré à la troisième octave. Seuls deux types d'accords sont ici supportés (majeur et mineur), et le modèle de Markov caché présenté comporte donc 24 états (correspondants aux 24 accords majeurs et mineurs différents). La méthode atteint une précision satisfaisante (plus de 70% d'accords correctement estimés), et commet une majorité d'erreurs de mode (un Do Majeur est estimé à la place d'un Do mineur), et de dominante (un Do Majeur est estimé à la place d'un Fa Majeur).

En 2009, Weller et al. [WEJ09] et Reed et al. [RUS+09] proposent des améliorations dans les modèles statistiques de reconnaissance d'accords, que ce soit par la prise en compte des minima d'erreurs de classification ou dans l'approche des machines à vecteurs de support (*Support Vector Machine, SVM*).

En plus des paramètres statistiques, certains chercheurs ont cherché à incorporer des informations musicales dans ces modèles statistiques, aboutissant ainsi à des modèles hybrides. L'intérêt d'une telle prise en compte réside dans le fait que les informations musicales ne dépendent pas de bases de données d'entraînements, contrairement aux informations statistiques.

4.2.2.3 Modèles musicaux / hybrides

Les modèles hybrides, où une information musicale est apportée en plus du modèle statistique, représentent une part importante des travaux récents en reconnaissance d'accords depuis le signal audio. En 2005, Bello et Pickens [BP05] proposent ainsi une méthode très similaire à celle de Sheh et Ellis [SE03], mais incorporent dans la matrice de transition de leur modèle une pré-connaissance musicale : certaines transitions entre accords sont plus probables que d'autres, selon la théorie de la musique. Ils proposent ainsi de donner une plus grande probabilité de transition décroissante entre deux accords en fonction de la distance entre leurs fondamentales dans le cycle des quintes.

En 2008, Papadopoulos et Peeters [PP08] proposent un modèle, toujours fondé sur un HMM, utilisant des informations relatives à la pulsation pour affiner l'estimation d'accords. Leur système établit ainsi une reconnaissance d'accords (majeurs et mineurs) sur des durées de 1 temps ou 1 demi-temps. Ils montrent ainsi que la prise en compte des demi-temps leur permet d'augmenter la précision de leur système, en donnant de meilleurs résultats sur les morceaux où les changements d'accords n'ont pas lieu sur les temps.

La même année, Lee et Slaney [LS07, LS08] proposent un autre modèle, utilisant toujours un HMM, et incorporant des informations relatives à la tonalité générale de la musique. 24 HMM sont en réalité utilisés, chacun associé à une des 24 tonalités générales possibles. A chaque fois, 36 états, correspondant aux 12 accords majeurs, mineurs, et diminuées sont considérées. Cette méthode propose ainsi une estimation de la séquence d'accords et la tonalité générale de l'extrait analysé.

Plus récemment, Mauch [MD10] utilise un modèle bayésien dynamique combinant deux types de vecteurs chroma (grave et aigu), une estimation de pulsation et de tonalité. Mauch propose également d'incorporer la structure musicale dans le traitement du chromagramme [MND09]. Il fait l'hypothèse que les séquences d'accords sont inchangés pour un même motif musical répété (couplet ou refrain). Il propose donc de moyenner les parties correspondant à des sections musicales répétées. Ainsi, si un couplet est répété trois fois, le premier vecteur chroma de ces trois couplets sera, après transformation, identique et égal à la moyenne des trois vecteurs chroma commençant les trois couplets avant transformation. Il montre ainsi que certaines erreurs peuvent être corrigées, car dues à des phénomènes qui ne se répètent pas d'une instance d'un couplet ou d'un refrain à une autre (notes de passage ou paroles différentes, par exemple).

La même année, Pauwels [PM10], propose également un modèle probabiliste utilisant une information musicale : la distance de Lerdahl. Cette distance est adaptée pour quantifier les probabilités de transitions entre deux accords, de manière à retrouver la séquence d'accords maximisant la probabilité d'occurrence associée. Pauwels note que l'intégration de cette information musicale, dans le système proposé, ne conduit pas à une amélioration très importante (avec un gain en précision inférieur à 2%). Selon lui, cette information corrobore le fait qu'une méthode ignorant toute considération musicale peut se révéler plus performante que des méthodes utilisant une telle information.

Une nouvelle méthode pour l'estimation des accords depuis le signal audio sera proposée dans le chapitre 5. Cette méthode permet une estimation simultanée des accords et des tonalités locales, et comporte une phase de recherche de modèles d'accords, une mesure de distance musicale et la recherche d'un chemin optimal parmi différents états possibles, la plaçant ainsi dans la troisième catégorie de méthodes décrites ci-dessus.

4.3 Tonalités

Cette section présente les méthodes existantes pour l'estimation de tonalité. De nouvelles méthodes, constituant des contributions, seront également proposées pour l'estimation de tonalité depuis une représentation symbolique de la musique. Parmi ces contributions, nous proposerons une méthode d'estimation de tonalité en utilisant des informations issues des séquences d'accords. Ceci nous permettra de vérifier si l'hypothèse émise précédemment, en mettant en évidence une amélioration de l'estimation des accords en utilisant les propriétés des groupes de notes, se trouve également vérifiée pour l'estimation de tonalité.

4.3.1 Depuis une représentation symbolique de la musique

Deux des premières méthodes de reconnaissance de tonalité furent proposées par Longuet-Higgins et Steedman en 1971 [LHS71], et Holtzman en 1977 [Hol77]. L'algorithme de Longuet-Higgins et Steedman repose sur une élimination progressive des tonalités candidates, en observant si les notes de l'extrait étudié appartiennent aux gammes des tonalités considérées. Si toutes les tonalités candidates sont éliminées, ou que l'analyse s'achève avec plusieurs candidates restantes, la méthode observe la première note de l'extrait, et déclare qu'il s'agit de la tonique de la pièce de musique, définissant ainsi la tonalité. Cette méthode ne s'applique qu'aux mélodies, et souffre d'un inconvénient majeur : une seule note étrangère à la gamme associée à la tonalité du morceau élimine cette tonalité des tonalités candidates. La méthode de Holtzman utilise elle aussi une série de règles, mais appliquées cette fois aux relations entre toniques et tierces, toniques et quintes, et triades présentes dans le morceau. Cette méthode considère seulement le début et la fin des extraits analysés. Comme souligné ensuite par Temperley [Tem99b], cette restriction est problématique, notamment dans le cas où les mélodies sont longues, et où la tonalité n'est pas nécessairement affirmée au début et à la fin des extraits considérés.

En 1990 Krumhansl [Kru90], à la suite des expériences psychoacoustiques [KK82], propose un autre algorithme, inspiré des méthodes de Longuet-Higgins et Holtzman, et utilise profils de notes (présentés section 2.2.1.2), et profils de tonalité (présentés section 3.1.2). Cet algorithme, très populaire, est souvent désigné sous le nom d'algorithme KS.

Une fois les profils de tonalité obtenus, l'algorithme les compare un à un au profil de la pièce étudiée. Comme il existe 12 profils majeurs et autant de profils mineurs, 24 corrélations sont calculées pour estimer le degré de ressemblance entre ces profils. Plus le profil de notes est similaire à un profil de tonalité, plus la corrélation entre ces deux profils est élevée. Il suffit alors de considérer le profil de tonalité offrant la plus grande corrélation avec le profil de notes pour pouvoir estimer sa tonalité. Le calcul de corrélation proposé par Krumhansl [Kru90] est le suivant :

$$C(n,r) = \frac{\sum_{i=0}^{11}(n(i)-\bar{n})(r(i)-\bar{r})}{\sqrt{\sum_{i=0}^{11}(n(i)-\bar{n})^2(r(i)-\bar{r})^2}}$$

avec n le profil de notes, r le profil de tonalité, $n(i)$ la i-ème composante du profil de notes, $r(i)$ la i-ème composante du profil de tonalité, et $\bar{n}$ et $\bar{r}$ les moyennes respectives des deux profils.
L'algorithme KS, évaluée par Krumhansl sur 48 pièces de musique classique, estime correctement la tonalité dans 44 cas, soit une précision supérieure à 90%.

Temperley [Tem99b] suggère par la suite une modification du calcul de corrélation, en adoptant un simple produit scalaire :

$$C_(n,r) = \sum_{i=0}^{11} n(i)r(i)$$

Temperley propose également une modification des profils de tonalité utilisés pour la comparaison avec le profil de notes de la pièce étudiée. Ces modifications ont été présentés dans le chapitre précédent, dans la section 3.1.2. De telles modifications permettent, sur une étude d'extraits de la suite pour violoncelle n° 3 de J.S. Bach, de diminuer par deux les erreurs commises par l'algorithme KS [Tem99a].

En 2007, Madsen et Widmer [MW07] adoptent une approche quelque peu différente. En utilisant les observations de Li et Huron [LH06], ils tiennent compte de l'ordre d'apparition des notes, en plus de leurs fréquences. Deux distributions de notes identiques peuvent en effet accepter deux séquences de notes différentes, qui peuvent donc potentiellement induire deux tonalités différentes. La figure 49 illustre l'influence de l'ordre d'apparition des notes. Sur l'extrait de la fugue de Bach, les La sont à contre temps, alors que la descente Sol-Fa-Mi-Ré se situe sur les temps, accentue le Ré final et affirme la tonalité de Ré mineur. A l'inverse, l'extrait proposé en dessous présente les différents La sur les temps forts (premiers et quatrièmes temps), et termine sur un enchaînement La-Mi-La, affirmant une relation de dominante et de tonique pour le Mi et le La, affirmant ainsi une tonalité de La mineur.
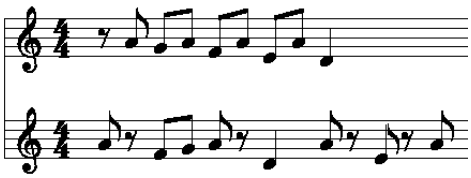


FIGURE 49 – En haut figure un extrait en Ré mineur (extrait de la célèbre fugue de J.S. Bach). En bas, nous proposons un extrait composé des mêmes notes, avec les mêmes durées, mais dans un ordre différent. Ce dernier semble plutôt être en La mineur.

Suite à de telles considérations, Madsen et Widmer ont adopté une représentation matricielle, pour dénombrer non plus les notes, mais les intervalles entre les notes. Chacun des 12x12=144 intervalles possibles se voit donc attribuer une probabilité d'apparition, représentée dans une matrice 12x12 (nommée *interval profile*, voir figure 50). Comme pour les profils de Krumhansl et Temperley, la matrice d'intervalles d'une pièce de musique est comparée aux matrices de tonalité, et un calcul de corrélation est mené pour estimer la tonalité :

$$C_{interval}(ref, piece) = \sum_{i=1}^{12}\sum_{j=1}^{12} M_{ref}[i,j]M_{piece}[i,j]$$

où $C_{interval}$ est la corrélation entre les deux profils d'intervalles, M_{piece} le profil d'intervalles de la pièce et M_{ref} le profil de tonalité, ces deux profils étant matriciels.

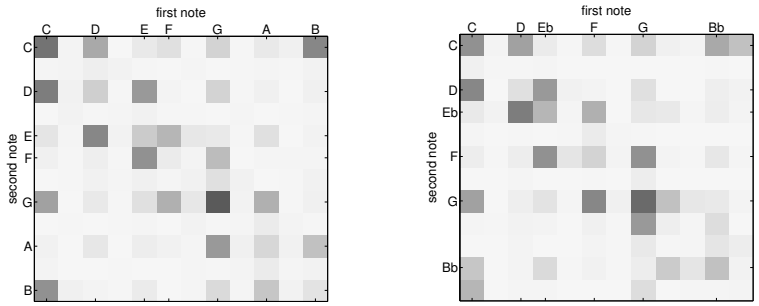


FIGURE 50 – Les profils d'intervalles de Madsen (majeur à gauche, mineur à droite), construits à partir des chorals de J.S. Bach (communication personnelle).

Madsen propose une évaluation de ces méthodes d'analyse de tonalité. Pour ce faire, une base de données, regroupant plusieurs morceaux dont la tonalité est annotée, est constituée. Les méthodes évaluées sont ensuite appliquées sur l'intégralité des morceaux de la base de données, et le score d'une méthode est obtenu par la formule :

$$\text{Score} = \frac{\text{Nombre de morceaux dont la tonalité est correctement estimée}}{\text{Nombre de morceaux de la base de données}}$$

Les évaluations de cette méthode matricielle se sont avérées tout à fait comparables aux évaluations des méthodes utilisant l'algorithme KS modifié par Temperley. Sur la base de données utilisée par Madsen [ET04], les deux méthodes obtiennent des scores avoisinant les 75%. En revanche, elles donnent certains résultats différents sur certaines pièces de musique (la méthode matricielle peut échouer là où la méthode vectorielle donne le bon résultat, et vice versa). Madsen a ainsi constaté que dans presque 85% des cas, au moins une des deux méthodes donne le bon résultat. Puisque les informations issues des ces deux analyses semblent être différentes, il semble judicieux de tenter de les combiner pour améliorer les résultats.

4.3.1.1 Une première proposition d'amélioration pour les méthodes d'analyse de tonalité

En partant des observations précédentes, nous proposons une nouvelle méthode d'évaluation de la tonalité [RRH08]. Cette méthode est fondée sur la combinaison des profils vectoriels et des profils matriciels. Il a ainsi été choisi de sommer les corrélations entre les profils (de notes et d'intervalles) de la pièce de musique et les profils (de notes et d'intervalles) de tonalité. Ces corrélations ont été normalisées, afin que chaque méthode ait le même apport en information :

$$C = C_{pitch} + C_{interval}$$

où C est la corrélation finale, C_{pitch} la corrélation entre les profils de notes et $C_{interval}$ la corrélation entre les profils d'intervalles

Bases de données

Deux bases de données ont été utilisées pour évaluer la nouvelle méthode proposée. La première est la base d'entraînement du MIREX (*Music Information Retrieval Evaluation eXchange*) qui est un ensemble de concours annuels, dont l'objectif est de comparer différentes méthodes proposées par la communauté scientifique pour l'extraction de données musicales diverses [Dow08, DEBJ10]. Pour l'évaluation de la tonalité en particulier, le MIREX avait proposé une base d'entraînement, contenant 100 extraits de musique classique, tirés de Classical Music Archives[3], et dont les tonalités sont connues (voir annexe A.2.1). La seconde base de données est un regroupement de chansons traditionnelles finlandaises [ET04], qui contient plus de 8000 mélodies traditionnelles transcrites par la société littéraire finlandaise (SKS). Tous les fichiers de cette base de données ont été annotés, de

3. `http://www.classicalarchives.com`

telle sorte que leurs tonalités sont également connues (voir annexe A.2.2)
Musicalement parlant, les deux bases de données sont relativement différentes, puisqu'il ne s'agit pas de musique classique, mais de musique traditionnelle pour les chansons finlandaises. De plus, ces dernières étant exclusivement composées de mélodies, elles sont monophoniques alors que les extraits classiques de la base d'entraînement du MIREX sont polyphoniques.

Protocole d'évaluation

Le protocole d'évaluation est le suivant. Pour chaque base de données, chaque morceau est analysé selon :

- la méthode des profils de notes classique en utilisant les profils de tonalité de Temperley,
- la méthode des intervalles proposée par Madsen [4],
- la combinaison des deux méthodes que nous proposons.

Pour chacun des morceaux et pour chacune des méthodes, un score est attribué :

- la tonalité estimée est exactement la tonalité de la vérité terrain. Dans ce cas, la tonalité est déclarée *correcte*, et se voit attribuer un score de 1 pour le calcul du score MIREX,
- la tonalité estimée est voisine de la vérité terrain (par exemple, on détecte Do Majeur alors que la vérité terrain stipule Fa Majeur). Dans ce cas, la tonalité est déclarée *voisine* et se voit attribuer un score de 0.5 pour le calcul du score MIREX,
- la tonalité estimée de la vérité terrain (par exemple, on détecte Do Majeur alors que la vérité terrain stipule La mineur). Dans ce cas, la tonalité est déclarée *relative* et se voit attribuer un score de 0.3 pour le calcul du score MIREX,
- sinon, la tonalité est déclarée *autre*, et se voit attribuer un score de 0 pour le calcul du score MIREX.

Les scores mentionnés sont ceux proposés par le MIREX pour l'évaluation de la tonalité.

Résultats

On trouvera dans le tableau 12 les résultats concernant l'évaluation de la méthode des profils de notes classique, la méthode en intervalles présentée par Madsen et la combinaison que nous proposons. Les trois méthodes ont été évaluées sur les deux bases de données présentées précédemment.

On note que la méthode proposée (*hybride*) permet une amélioration des résultats, en augmentant le score MIREX dans les deux bases de données (base d'entraînement du

4. nous tenons à remercier Soren Madsen pour sa communication des profils en intervalles appris sur les chorals de Bach, sans lesquels l'évaluation de cette méthode n'aurait pas été possible

Méthode	BD	Tonalité	Erreurs			Score MIREX
		correcte	relatif	voisin	autres	
PP	M	67.4	14.7	17.9	0.0	80.8
Intervalles	M	83.2	4.2	11.6	1.1	90.2
Hybride	M	**83.2**	4.2	12.6	**0.0**	**90.8**
PP	FF	71.0	7.0	16.5	5.5	81.4
Intervalles	FF	72.4	7.0	15.6	5.0	82.3
Hybride	FF	**75.7**	4.7	15.7	**3.9**	**85.0**

Table 12 – Précision de l'estimation de tonalité selon la méthode étudiée. Les résultats présentés sont en pourcentage du nombre de fichiers de la base de données concernée, pour la méthode concernée. *PP* désigne la méthode des profils de notes classique, avec les profils de tonalité de Temperley, *Intervalles* désigne la méthode utilisant les intervalles présentée par Madsen, et *Hybride* désigne notre proposition de combinaison des deux méthodes. *M* désigne la base d'entraînement du MIREX, alors que *FF* désigne la base de données des chansons traditionnelles finlandaises.

MIREX et chansons traditionnelles finlandaises). Si sur la base d'entraînement du MIREX (constituée d'une centaine de morceaux uniquement), la méthode proposée a permis de détecter le même nombre de tonalités correctes que la méthode des intervalles, sur la base des chansons finlandaises, elle a permis de calculer la tonalité indiquée pour 264 morceaux de plus que la méthode matricielle et 374 morceaux de plus que la méthode des profils de notes classique. On remarque en outre que l'amélioration du score MIREX découle majoritairement de la diminution du nombre d'erreurs lointaines commises (c'est-à-dire les erreurs n'étant ni de type ton relatif, ni de type ton voisin). Les tonalités estimées par la méthode proposée peuvent donc faire l'objet d'une correction.

Corrections de la méthode hybride

D'après les observations précédentes, une correction de la méthode hybride est envisageable en transformant certaines erreurs de tons voisins ou relatifs en tonalité correcte.

Une première idée est de mettre au point une méthode permettant de déterminer si le morceau étudié est majeur ou mineur (et ainsi, pouvoir décider entre deux tonalités relatives, qui ont par définition deux modes opposés). Nous avons donc construit deux profils spécifiques à la résolution de ce problème (voir figure 51).

Le calcul de corrélation entre le profil de la pièce et ces profils spécifiques (adaptés à la tonalité relative de la tonalité proposée) peut permettre de corriger la tonalité proposée. Ainsi, si la méthode précédente propose une tonalité de Do Majeur, mais que la corrélation entre le profil de la pièce et le profil spécifique de La mineur est supérieure à la corrélation

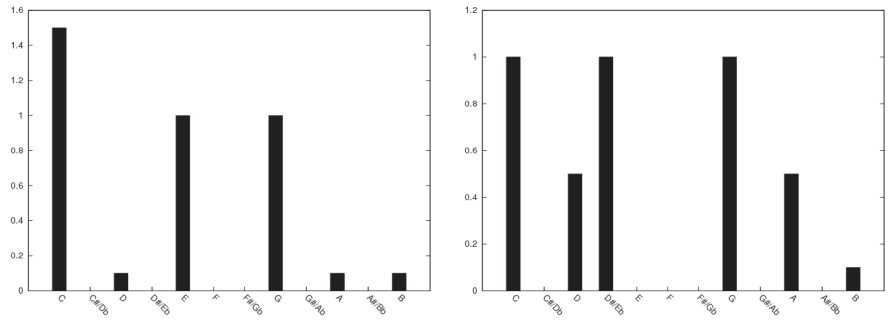


FIGURE 51 – Les deux profils permettant de connaître le mode de la pièce étudiée (majeur à gauche, mineur à droite).

BD	Corrections	Tonalité correcte	Erreurs			Score MIREX
			relatif	voisin	autre	
M	non	83.2	4.2	12.6	0.0	90.8
M	oui	**89.5**	4.2	6.3	**0.0**	**93.9**
FF	non	75.7	4.7	15.7	3.9	85.0
FF	oui	**76.8**	4.7	14.5	**3.9**	**85.6**

TABLE 13 – Corrections des résultats, sur les deux bases de données (*M* pour base d'entraînement MIREX, *FF* pour la base de données des chansons traditionnelles finlandaises), en pourcentage du nombre de fichiers pour la base de données concernée.

entre le profil de la pièce et le profil spécifique de Do Majeur, il est alors possible de procéder à la correction et de proposer non plus Do Majeur mais La mineur.

Un autre moyen de pouvoir déterminer si un morceau est majeur ou mineur peut être d'observer la répartition des intervalles. On peut en effet affirmer qu'en Do Majeur par exemple, il existe beaucoup plus d'intervalles (do - mi) que d'intervalles (do - mi♭) (comme on peut le voir sur la figure 50). De plus, si une erreur de ton relatif se produit avec une tonalité détectée majeure, l'intervalle formé par la tonique de la tonalité mineure relative et sa sensible peut apparaître. Il est alors possible de corriger en relatif mineur une estimation de tonalité majeure qui présente :

- un grand nombre de tierces mineures à partir de sa tonique, comparativement aux nombre de tierces majeures à partir de cette même note,
- un nombre important d'intervalles (dominante augmentée - sus dominante), que l'on peut donc voir comme un intervalle (sensible majeure - tonique) dans le ton relatif mineur.

Le tableau 13 présente les résultats de la correction.

On observe que la correction a pu encore améliorer les résultats, notamment sur la base d'entraînement MIREX. Certaines erreurs ont donc pu être corrigées.

Conclusion

Nous avons montré qu'une méthode d'estimation de tonalité utilisant des profils vectoriels et matriciels peut bénéficier de la complémentarité des informations apportées par ces deux types de profils. Ce résultat a fait l'objet d'une publication [RRH08].

Néanmoins, pour cette méthode comme pour les autres méthodes d'estimation de tonalité existante, l'information musicale prise en compte provient invariablement des notes de musique, qui constituent un niveau de description bruité de l'harmonie : toutes les notes ne participent en effet pas à l'harmonie, et peuvent constituer des ornements. Il serait donc intéressant d'étudier la prise en compte d'une information tonale moins bruitée que les notes pour l'estimation de tonalité.

4.3.1.2 Utilisation des accords pour l'analyse de tonalité

Nous proposons dans cette section d'utiliser l'information apportée par la séquence d'accords d'un morceau de musique pour l'analyse de sa tonalité. Nous considérons des morceaux dont la suite d'accords est connue et qui constitue l'entrée musicale pour l'estimation de la tonalité du morceau étudié. Nous proposons ainsi une nouvelle méthode utilisant une comparaison entre un profil de notes et des profils de tonalité. Les profils de notes ne sont en revanche plus construits à partir des notes, mais à partir de la séquence d'accords : pour chaque accord, la triade de notes correspondante est générée et ajoutée au profil de notes (la fondamentale se voyant attribuer un poids deux fois plus important que la tierce et la quinte). L'algorithme suivant décrit la construction du profil de notes à partir de la séquence d'accords :

```
initialiser le profil de tonalite
pour toute fondamentale x des accords
   generer les notes x (deux fois), tierce(x) et quinte(x)
   ajouter ces 3 notes au profil de tonalite
fin pour tout
retourner la tonalite correspondant au profil de reference maximisant
la correlation avec le profil de tonalite
```

Le calcul du maximum de corrélation est ensuite identique à l'algorithme KS, en utilisant le produit scalaire et les profils de tonalités proposés par Temperley (voir figure 27).

Exemple

Considérons les premières mesures de *Little Wing* de Jimi Hendrix (voir figure 52). Il est possible de construire un profil de notes de cet extrait selon la méthode classique [Kru90], pour lequel chaque note aura une contribution proportionnelle à sa durée.

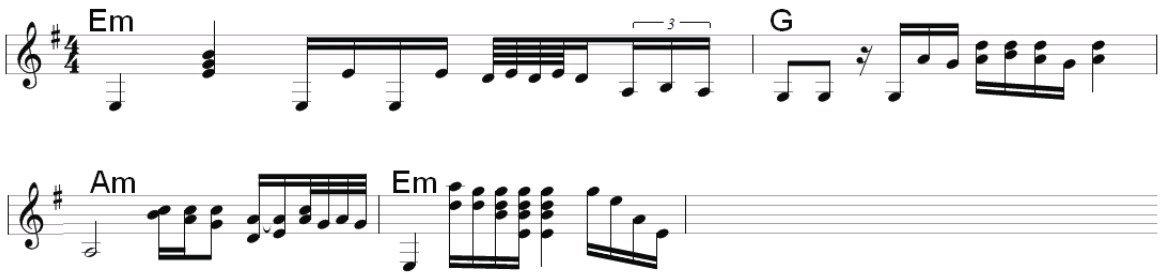


FIGURE 52 – Les quatre premières mesures de *Little Wing*, de Jimi Hendrix. Les accords sont notés au dessus de chaque mesure. L'extrait est en Mi mineur.

La méthode que nous proposons utilise un profil ne correspondant non plus à la fréquence de toutes les notes de l'extrait, mais des notes des accords de l'extrait. Ainsi, cet extrait comportant 4 accords (Emin, Gmaj, Amin, Emin), les notes ajoutées aux profils seront les suivantes (avec leur poids) :

- Mi(2) Sol(1) Si(1) (correspondant au Emin),
- Sol(2) Si(1) Ré(1) (correspondant au Gmaj),
- La(2) Do(1) Mi(1) (correspondant au Amin),
- Mi(2) Sol(1) Si(1) (correspondant au Emin).

Les profils de notes issus des deux méthodes sont différents, comme le montre la figure 53.

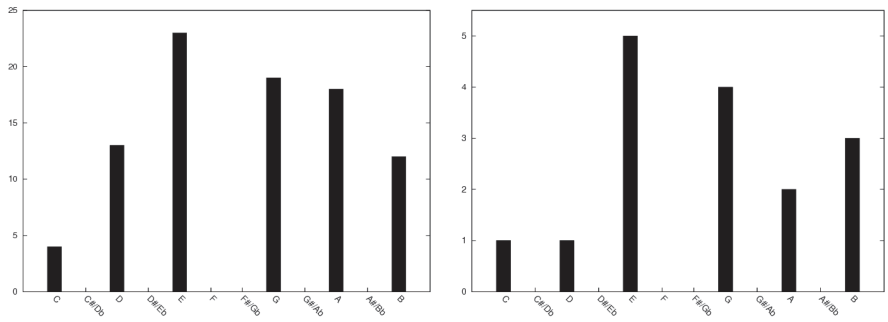


FIGURE 53 – Profils de notes des premières mesures de *Little Wing* de Jimi Hendrix (en Mi mineur), en utilisant la méthode classique (à gauche) et avec la méthode proposée (à droite). Les poids utilisés pour la méthode classique sont les suivants : 1 pour une double croche, 2 pour une croche, 4 pour une noire, etc.

On remarque par exemple que le Mi se voit conférer une plus grande importance avec le profil issu de la méthode proposée, tout comme le Sol et le Si, qui sont les autres notes de la triade correspondant à Mi mineur (Emin). Le profil issu de la méthode des profils de notes classique marque moins cette différence entre les notes de la triade et les autres notes, et apparaît donc comme plus bruité, notamment en raison du poids important conféré au La, qui se retrouve quasiment aussi important que le Sol et bien plus que le Si. On peut ainsi avoir l'intuition, en observant ces deux profils, que la tonalité correcte sera plus facilement estimée en utilisant le profil issu de la méthode proposée. Nous nous proposons donc de calculer les 24 corrélations avec les profils de tonalités pour ces deux profils, afin de vérifier cette hypothèse.

La figure 54 représente le résultat des corrélations entre les deux profils de notes différents (qui ont été normalisés) et les 24 profils de tonalité de Temperley.

On remarque que le profil de notes issu de la méthode proposée présente une corrélation maximale avec le profil de tonalité de Mi mineur, qui est la tonalité correcte de l'extrait musical. En revanche, le profil classique maximise la corrélation pour la tonalité de Sol Majeur (qui est le ton relatif de Mi mineur). On remarque également que Do Majeur présente une corrélation supérieure à Mi mineur, et donc que la tonalité correcte ne se retrouve qu'en troisième position en utilisant le profil classique. Dans cet exemple, la prise en compte des accords pour construire le profil de tonalité d'un morceau semble donc avoir pour effet de filtrer le bruit harmonique causé par la prise en compte de chacune des notes de la pièce étudiée.

Ce simple exemple, permet d'avoir un premier aperçu du fonctionnement de la méthode proposée, ainsi que de son efficacité supposée. Dans la section suivante, une évaluation de cette méthode est présentée, ainsi que sa comparaison avec la méthode des profils de notes classique.

Evaluation

Afin d'estimer une tonalité grâce à des séquences d'accords, il est nécessaire de disposer d'annotations des accords sur un nombre suffisamment important de morceaux de musique. Nous avons eu recours, comme en 4.2.1, aux transcriptions des séquences d'accords des Beatles par Christopher Harte. Néanmoins, ce dernier n'a pas pu fournir d'annotations sur la tonalité de chacun des morceaux. Nous avons donc eu recours aux travaux d'Allan Pollack [5]. Ce dernier a réalisé une étude de la structure des morceaux des *Beatles*, en mentionnant en particulier pour chacun la tonalité.

5. `http://www.icce.rug.nl/\~soundscapes/DATABASES/AWP/awp-alphabet.shtml`

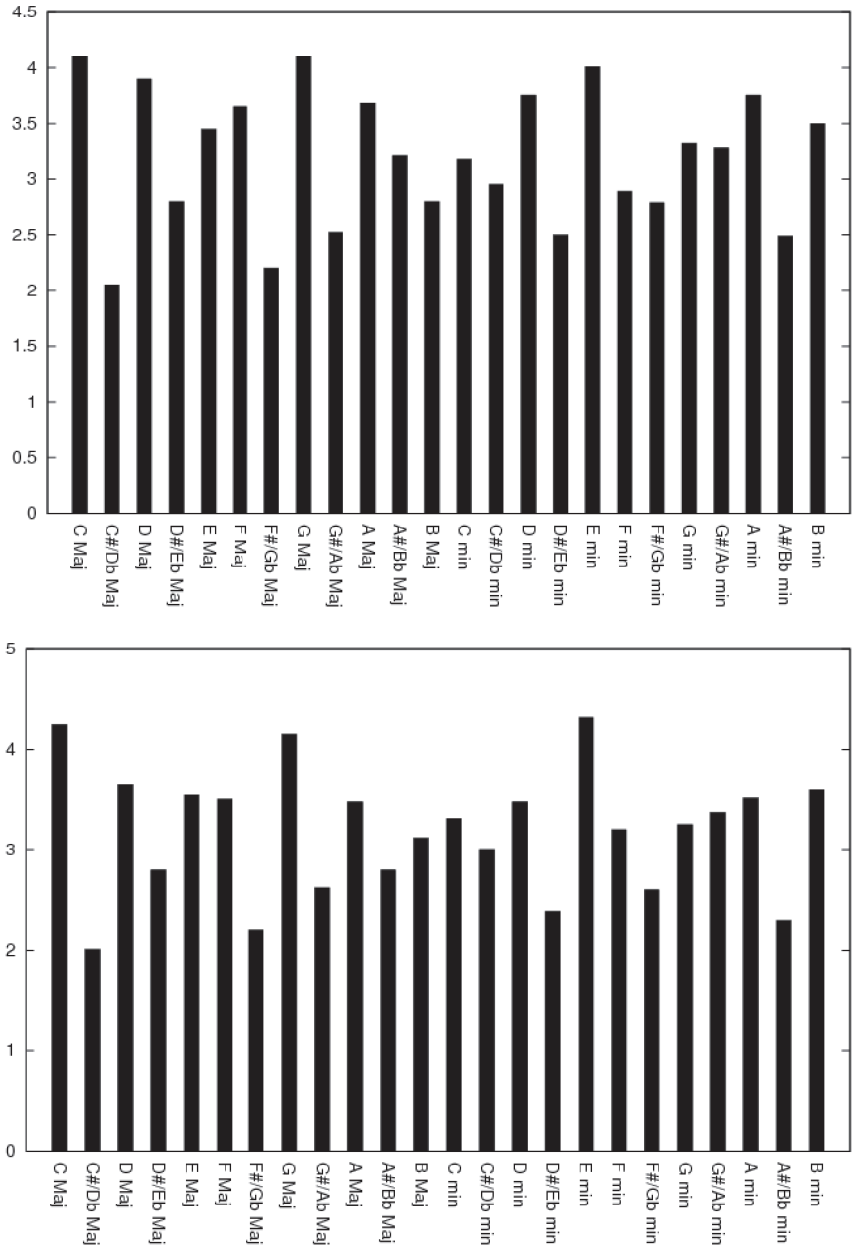


FIGURE 54 – Résultats des corrélations (par produit scalaire) entre les 24 profils de tonalité et les profils de notes de l'extrait de *Little Wing* obtenus par la méthode des profils de notes classique (en haut) et la méthode proposé (en bas). Les tonalités majeures apparaissent en majuscule (à gauche) et les tonalités mineures en minuscule (à droite). La tonalité de l'extrait (Mi mineur) apparaît présente la plus forte corrélation avec la méthode proposée, ce qui n'est pas le cas avec la méthode classique.

Bases de données

Les morceaux sélectionnés pour l'évaluation sont donc les morceaux des Beatles dont :
- nous disposions de la transcription accord par accord par Christopher Harte,
- nous disposions d'une tonalité satisfaisante par Allan Pollack (dans ses notes, Allan Pollack cite notamment des morceaux dont la tonalité apparaît comme ambiguë, comme par exemple *Birthday*, où deux tonalités différentes sont mentionnées),
- nous disposions d'une représentation symbolique (fichier MIDI).

Ainsi, 86 morceaux ont finalement été retenus pour l'évaluation (voir annexe A.2.3).

Protocole d'évaluation

Pour chaque titre de la base de données, deux tonalités sont calculées :
- la première, selon la méthode d'analyse proposée : pour chaque accord de la vérité terrain, on génère la triade correspondante que l'on ajoute au profil de notes, lequel est ensuite comparé aux 24 profils de tonalité. Le profil de tonalité donnant la plus grande corrélation avec le profil de notes donne la tonalité estimée,
- la seconde, selon la méthode d'analyse classique : on procède de la même manière, mais le profil de notes n'est pas construit à partir des notes issues de la triade des accords, mais à partir de toutes les notes de la pièce de musique.

On compare ensuite les deux tonalités estimées selon ces deux méthodes à la vérité terrain mentionnée par Allan Pollack, en distinguant différents cas correspondant à l'attribution des différents scores MIREX (voir section 4.3.1.1).

Résultats

On trouvera les résultats de l'évaluation des accords dans le tableau 14. Les scores présentés sont en pourcentage du nombre total de morceaux (86). On trouvera en particulier le score MIREX, qui prend en compte les types d'erreurs rencontrés.

Méthode	Tonalité correcte	Erreurs				Score MIREX
		relatif	voisin	parallèle	autre	
PP notes	66.28	1.16	12.79	1.16	18.6	73.26
PP accords	**77.91**	0	11.62	0	**10.47**	**83.72**

Table 14 – Résultats de l'analyse des tonalités. *PP notes* désigne la méthode classique, en remplissant le profil du morceau avec les notes, alors que *PP accords* désigne la méthode proposée (pour laquelle le profil est construit à partir des accords).

La méthode proposée permet une amélioration de plus de 10% concernant la détection de tonalités correctes. Cette amélioration s'accompagne d'une forte diminution du nombre d'erreurs ne résultant pas de tons voisins ou de tons relatifs (diminution de plus de 8% pour notre méthode), qui implique que près de 90% des tonalités détectées sont soit correctes, soit voisines de la tonalité correcte, ce qui ouvre des perspectives quant à la correction de ces erreurs. Enfin, le score MIREX, qui mesure l'efficacité globale, est amélioré de plus de 10 points.

Conclusion

Nous avons étudié la prise en compte d'une information tonale moins bruitée que les notes de musique, ici la séquence d'accords, pour l'estimation de tonalité. Une telle considération permet une amélioration de la précision des tonalités estimées.
Dans la méthode présentée, les accords étaient néanmoins issus d'annotations. Il reste donc à déterminer si une telle amélioration peut toujours être atteinte en utilisant des séquences d'accords estimées, et non annotées. Cette question sera abordée dans le chapitre suivant.

4.3.2 Depuis le signal audio

Après avoir vu les méthodes d'estimation de tonalité depuis une représentation symbolique de la musique, nous étudions dans cette section les méthodes considérant le signal audio comme entrée musicale. Ces méthodes se divisent, dans la littérature, en deux grandes catégories :

- les méthodes utilisant une adaptation de l'algorithme KS (vu en section 4.3.1),
- les méthodes utilisant une approche utilisant des modèles statistiques, comme les HMM (présentés en 4.2.2.2).

4.3.2.1 Adaptations de l'algorithme KS

Une des premières adaptations de l'algorithme KS au signal audio fut menée en 2004 par Gomez et Herrera [GH04], qui comparent l'adaptation directe de cette méthode en utilisant des vecteurs chroma HPCP (voir section 2.2.2.1) moyennés sur les trames constituant une pièce de musique. Ce vecteur moyen joue ensuite le rôle du profil de notes dans l'algorithme KS, et est donc comparé à des profils de tonalité (identiques à la méthode proposée en 1990 par Krumhansl). La même année, Pauws propose [Pau04] une approche très similaire, en utilisant cette fois une pondération des vecteurs chroma, et en ne sélectionnant que certaines parties de l'extrait (typiquement le début et la fin) pour construire son vecteur chroma moyen, avant de le comparer aux profils de tonalités.

En 2005, Izmirli [Izm05] propose deux méthodes similaires à l'algorithme KS, l'une utilisant des modèles spectraux et l'autre des vecteurs chroma. Pour chacune de ces méthodes, des profils de tonalités sont créés à partir d'enregistrements de piano annotés, et comparés aux profils de tonalités de Krumhansl [KK82] et de Temperley [Tem99a]. Cette étude montre notamment que les profils de tonalités proposés par Temperley semblent plus efficaces dans le cadre de l'algorithme KS. En 2006, Izmirli propose une amélioration possible de l'algorithme KS en utilisant un chroma moyen, à la manière de [GH04], mais construit cette fois en considérant des groupes de vecteurs chroma de taille variable, en cherchant à maximiser la corrélation avec les profils de tonalités [Izm06].

4.3.2.2 Méthodes utilisant une approche statistique

Parmi les méthodes d'estimation de tonalité depuis le signal audio utilisant une approche statistique, on citera la méthode proposée par Peeters en 2006 [Pee06], qui construit par apprentissage deux profils de tonalités en transformant tous les vecteurs chroma annotés de la base d'apprentissage en des profils de Do Majeur et Do mineur (par rotation d'indice vectoriel). La méthode construit ensuite, toujours par rotation d'indice, 24 HMM correspondant aux 24 tonalités différentes. Pour chaque morceau de musique étudié, la tonalité estimée correspond au maximum de probabilité de la séquence de vecteurs chroma selon les 24 HMM.
Une comparaison entre l'efficacité des profils appris et celle des profils de Temperley est ensuite présentée, comme cela avait été le cas dans [GH04]. Les conclusions sont identiques : les profils de Temperley donnent de meilleurs résultats que les profils appris statistiquement, sans aucune connaissance musicale préétablie.

Parmi ces méthodes statistiques, on trouve également des travaux se permettant d'identifier des modulations, et d'estimer des tonalités locales depuis le signal audio. Harte [HSG06] et Chai[CV07] ont ainsi proposé des méthodes reposant sur l'élaboration de HMM, et intégrant la probabilité pour chaque état d'être un point de modulation entre deux tonalités locales. En 2007, Izmirli [Izm07] propose un modèle utilisant une matrice de factorisation non-négative pour la segmentation des tonalités locales. Des groupes de vecteurs chroma sont donc successivement considérés comme étant candidats à la constitution d'un segment musical sans modulation. La taille de ces groupes de vecteurs chroma est fixée empiriquement.

Une nouvelle méthode d'estimation des tonalités locales est proposée dans le chapitre suivant.

Chapitre 5

Estimation conjointe des accords et des tonalités

Face à certaines limites des méthodes d'estimations d'accords et de tonalités existantes, en particulier une phase d'apprentissage nécessitant de grandes quantités de données et dont l'efficacité dépend des styles musicaux utilisés pour l'apprentissage, nous souhaitons proposer une nouvelle méthode d'estimation d'accords et de tonalités. Nous avons montré précédemment qu'il existait une interaction entre les différentes informations tonales : la séquence d'accords, par exemple, influence fortement la tonalité établie. Ainsi, estimer conjointement ces deux informations présente non seulement un sens musical, mais ouvre l'hypothèse d'un bénéfice mutuel pour l'estimation d'accords et de tonalités.
Ce chapitre présente ainsi une nouvelle méthode pour l'estimation conjointe des accords et des tonalités locales. Cette méthode peut être appliquée depuis une représentation symbolique de la musique ou depuis le signal audio. Un aperçu général de la méthode est présenté section 5.1, avant que les spécificités de l'application de cette méthode à une représentation symbolique et au signal audio ne soient développées sections 5.2 et 5.3. On trouvera également dans ces deux sections une évaluation de cette méthode.

5.1 Principe général

Avant de passer à la description complète de la méthode proposée depuis deux types d'entrées distincts (représentation symbolique et signal audio), nous proposons le principe général de cette méthode, et décrivons dans cette section les processus d'estimation communs à la musique symbolique et au signal audio. L'entrée musicale est tout d'abord segmentée, puis pour chaque segment, des candidats accords et tonalités sont énumérés, et rassemblés en couple *(accord, tonalité)*. Un graphe acyclique est ensuite créé, en reliant chaque couple associé à un segment de l'entrée à tous les couples du segment suivant.

Chaque arc ainsi formé est pondéré par une distance musicale estimant une mesure entre deux couples. Un processus de programmation dynamique permet ensuite de déterminer le chemin optimal dans ce graphe. La séquence d'accords et de tonalités locales finalement estimée par le système correspond à la séquence des couples *(accord, tonalité)* le long du chemin minimal.

5.1.1 Segmentation de l'entrée musicale

Le premier point de l'analyse du morceau consiste en une segmentation de l'entrée musicale (voir figure 55), chacun des segment ayant vocation finale à être étiqueté par un accord et une tonalité. Ce processus de segmentation diffère lorsqu'il s'agit de musique symbolique ou de signal audio. Depuis une représentation symbolique, un algorithme est proposé pour la création des segments, permettant ainsi de rassembler certaines notes jouées dans un même segment, ou au contraire de les séparer dans des segments différents. Depuis le signal audio, de même que dans la totalité des méthodes d'analyse existantes, l'entrée est découpée en trames. Comme nous le verrons plus loin (section 5.3), plusieurs résolutions différentes, correspondant à différents tailles de trames, pourront être considérées.
Cette segmentation de l'entrée musicale sert ensuite de support à l'énumération des candidats, puisque chaque segment donne lieu à une sélection de candidats accords et tonalités. Elle sert également d'unité de temps pour la séquence d'accords et de tonalités locales finalement générée par le système.

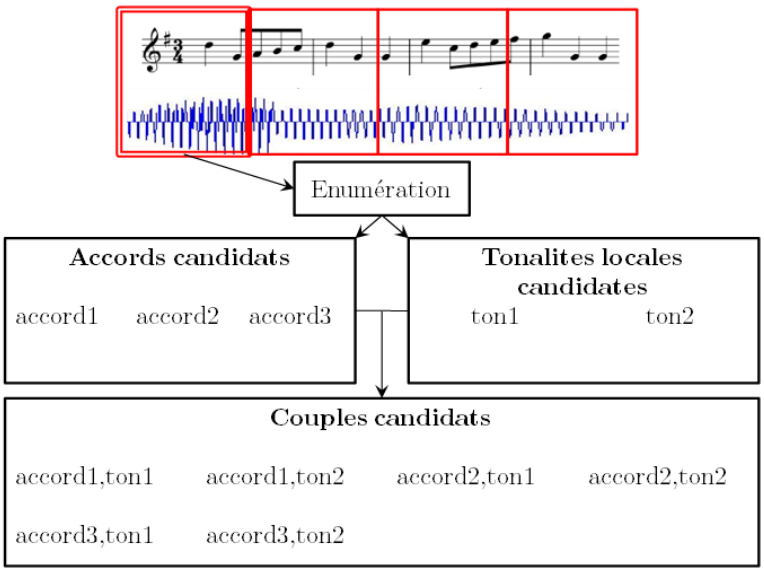


FIGURE 55 – Découpage de l'entrée musicale (symbolique ou audio) en segments et énumération de couples candidats.

5.1.2 Énumération de *couples candidats*

À partir de la segmentation de l'entrée musicale, différents accords et tonalités potentiels sont sélectionnés parmi un dictionnaire d'accords et de tonalités supportés par le système. Le processus d'énumération consiste à choisir, parmi les accords et les tonalités contenus dans le dictionnaire, certains accords et certaines tonalités possibles, qui constituent des candidats. Par accords candidats, on désigne ainsi les accords susceptibles d'être associés au segment étudié. Le but de cette phase d'énumération est ainsi de ne plus considérer tous les accords possibles issus du dictionnaire, mais de se limiter à quelques accords plausibles au vu du segment analysé. On désigne de même par tonalité candidate une tonalité plausible au vu du segment étudié. On souhaite ainsi énumérer l'accord et la tonalité correctes (telles qu'ils seraient annotés par un musicien) parmi les accords et tonalités candidats, pour que, lors de la sélection du chemin optimal, cet accord et cette tonalité soient choisis.
La manière avec laquelle nous nous proposons de mener cette énumération diffère significativement entre l'analyse depuis la musique symbolique et depuis le signal audio. Néanmoins, cette phase d'énumération conserve certaines propriétés communes, que ce soit lors de l'analyse depuis l'une ou l'autre de ces deux représentations :

- tous les accords et tonalités contenus dans le dictionnaire d'accords et de tonalités supportés sont successivement considérés,
- les accords et tonalités sélectionnées par l'énumération sont tous conservés pour la création du graphe (voir section 5.1.3), et les autres sont définitivement écartés,
- les n accords et m tonalités sélectionnés par l'énumération sont ensuite combinés en $c = n \times m$ couples, chaque couple étant de la forme *(accord, tonalité)*.

La figure 56 illustre le processus d'énumération des accords et des tonalités depuis un segment donné de l'entrée musicale. Dans la suite de ce chapitre, nous ferons référence aux couples *(accord, tonalité)* qui sont énumérés sous le nom de couples candidats.

5.1.3 Création d'un graphe de couples candidats

Une fois les couples candidats énumérés pour chaque segment de la pièce analysée, un graphe acyclique est créé, en reliant deux à deux tous les couples candidats d'un segment avec tous les couples candidats du segment suivant (voir figure 57). L'idée est ensuite de trouver un chemin dans ce graphe, conduisant à la meilleure séquence de couples candidats pour la pièce étudiée. Il existe une multitude de manières de considérer qu'un chemin dans un tel graphe est meilleur qu'un autre, comme il sera exposé dans la section 5.1.4.

Une distance musicale est utilisée pour permettre de pondérer la transition entre deux couples candidats. Plusieurs distances musicales, présentées en section 3.1.3 ont ainsi été

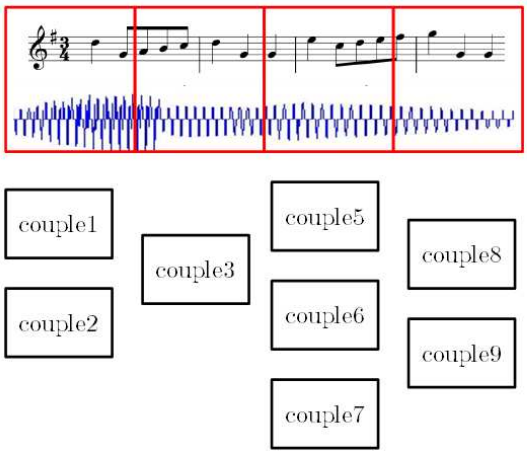


FIGURE 56 – Illustration des couples candidats pour les différents segments de la pièce analysée. Un nombre variable de couples candidats peut être observé sur les différents segments de la pièce.

étudiées pour cette pondération musicale. Plus la distance entre deux couples candidats $c_1 = (a_1, t_1)$ et $c_2 = (a_2, t_2)$, et moins il est probable d'observer l'accord a_1 dans la tonalité t_1 suivi de l'accord a_2 dans la tonalité t_2. Ainsi, plus la valeur, ou le coût, porté par un arc reliant c_1 à c_2 est élevé, plus il est musicalement improbable d'avoir la succession d'accords et de tonalités locales correspondant à la transition entre c_1 et c_2.

Tous les arcs du graphe sont ainsi pondérés par un coût musical entre les différents couples candidats.

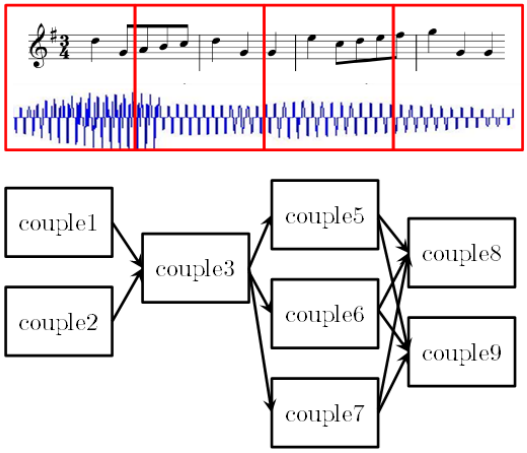


FIGURE 57 – Illustration du graphe acyclique de couples candidats.

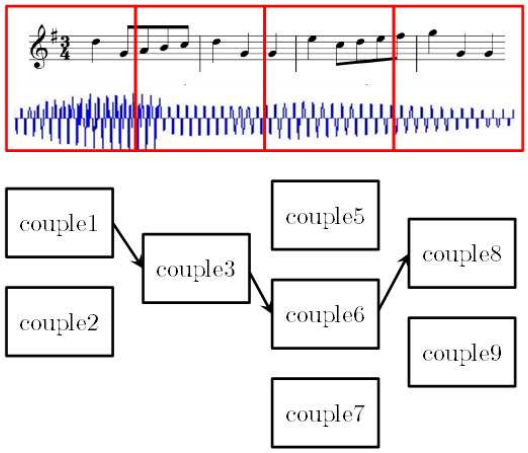


Figure 58 – Exemple de chemin optimal dans le graphe de couples candidats. Ce chemin donne la séquence finale d'accords et de tonalités locales.

5.1.4 Sélection du chemin optimal

Une fois le graphe de couples candidats établi, un processus doit permettre de sélectionner le chemin optimal dans ce graphe (voir figure 58) .

Cette tâche est résolue par programmation dynamique [Bel57]. Pour chaque candidat, un unique arc est choisi parmi les arcs incident. Cet arc optimise le chemin jusqu'à ce candidat. Une définition du chemin optimal doit ici être choisie. Nous qualifions de chemin optimal le chemin minimal, c'est-à-dire celui dont la somme des coûts des arcs le constituant est minimale.

Le problème appliqué au graphe entier se réduit donc à la recherche du minimum de :

$$\sum_{1 \leqslant i \leq n} \delta(c_i, c_{i+1})$$

où n est le nombre de segments de l'entrée musicale, et c_i et c_{i+1} désignent deux couples candidats appartenant à deux segments consécutifs, et donc reliés par un arc pondéré par la distance $\delta(c_i, c_{i+1})$.

L'arc incident choisi par la programmation dynamique est ainsi celui qui minimise la somme des coûts de transition le long du chemin menant à chaque couple candidat, comme illustré sur la figure 59(c). Il existe un chemin pour chaque couple candidat du dernier segment de la pièce étudiée. Parmi ces chemins, le chemin finalement retourné est le chemin dont la somme des coûts pour les arcs parcourus est minimale.

Le processus entier permettant la sélection du chemin optimal est illustré sur la figure 59.

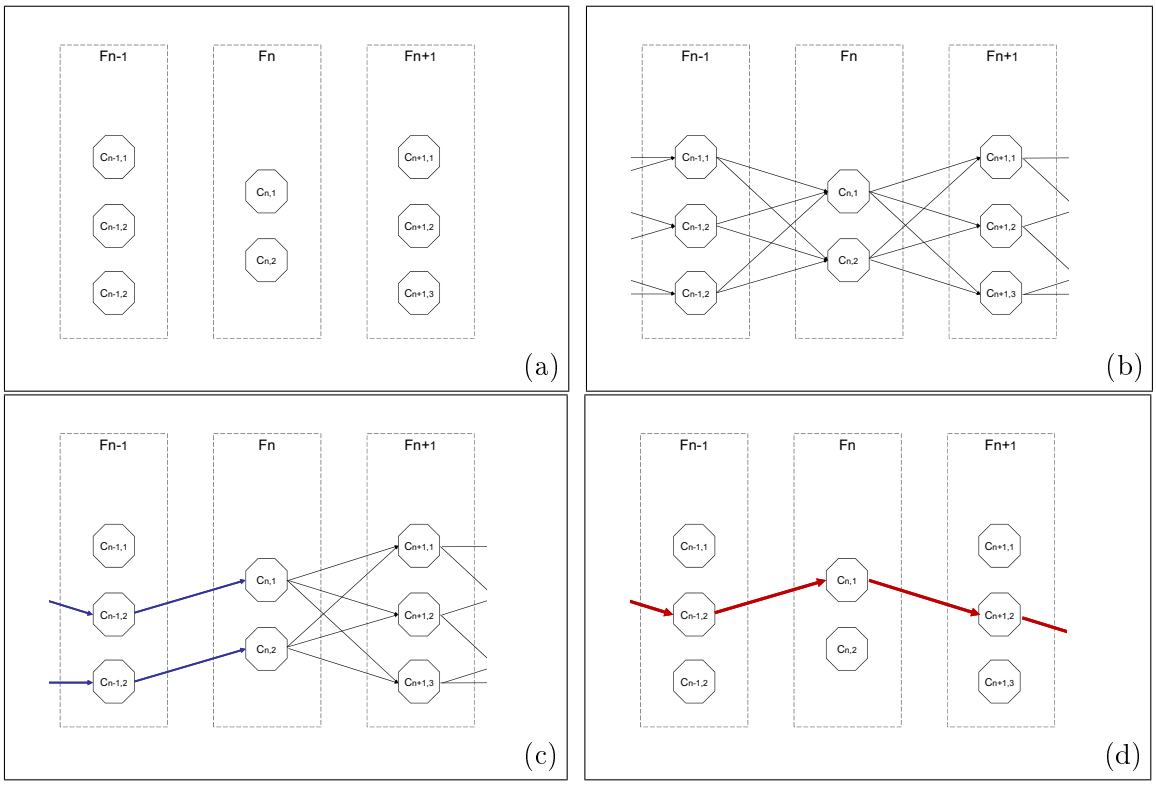


FIGURE 59 – (a) Énumération de candidats harmoniques $C_{i,j}$ pour des segments consécutifs F_i, ordonnés temporellement de gauche à droite. $C_{i,j}$ représente le $j$ème candidat harmonique pour le segment F_i. Le temps apparaît de gauche à droite. (b) Création des arcs pour former un graphe acyclique. Un arc relie chacun des candidats d'un segment donné à chacun des candidats du segment suivant. (c) La programmation dynamique consiste à choisir un unique chemin menant à chaque candidat d'un segment donné (ici, le segment n). (d) Sélection du chemin optimal. Les séquences d'accords et tonalités estimées sont ainsi constituées par la séquence des candidats le long du chemin optimal.

5.2 Application depuis une représentation symbolique de la musique

Nous présentons dans cette section l'application de la méthode générique décrite dans la section précédente dans le cas où l'entrée musicale analysée est sous forme symbolique (par exemple, sous la forme d'un fichier MIDI). La segmentation de l'entrée, les règles d'énumérations et les différentes distances musicales utilisables sont ainsi explicitées ci-après. Une évaluation de la méthode à travers ses différents paramètres sera ensuite proposée.

5.2.1 Segmentation

Nous proposons de mener la segmentation de la pièce de musique analysée en la découpant en groupes de notes (définis en 2.1.2). La justification principale est la suivante : un nouvel accord ne peut commencer que lorsque :

- une nouvelle classe de note est jouée,
- une classe de note cesse d'être jouée.

Il semble donc être raisonnable de se donner la possibilité de pouvoir changer d'accord (et de tonalité locale) dès qu'une nouvelle classe de note apparaît, ou qu'une classe de note n'est plus jouée. Chaque groupe de notes va ainsi donner lieu à une énumération d'accords candidats et de tonalités candidates.

5.2.2 Énumération des candidats

Qu'il s'agisse des accords ou tonalités locales, une approche fondée sur des règles (*rule based*) est utilisée pour déterminer quels candidats sont à considérer pour un groupe de notes donné.

5.2.2.1 Tonalités locales candidates

La méthode proposée pour déterminer si une tonalité est candidate pour un groupe de notes considéré est la suivante. Soit une tonalité t et sa gamme associée g, et un groupe de notes $gn = (n_1, ..., n_p)$, avec $1 \leq p \leq 12$. On déclare t candidate pour gn si et seulement si, chacune des classes de note n_i, $1 \leq i \leq p$, appartient à g. On considérera la gamme mélodique pour les tonalités mineures.
Ainsi, si le groupe de notes est (Do,Mi), les tonalités candidates énumérées seront Do Majeur, Fa Majeur, Sol Majeur, La mineur, Ré mineur et Mi mineur, car Do et Mi appartiennent tous les deux aux six gammes correspondant à ces six tonalités.

5.2.2.2 Accords candidats

Pour déterminer si un accord est candidat pour un groupe de notes donné, on utilise un ensemble de règles suivantes qui distinguent plusieurs types d'accords. Chaque règle est ainsi associé à la reconnaissance d'un type particulier d'accord, et vise à déterminer si un accord de ce type est identifiable dans le groupe de notes étudié. Si un accord est éligible selon une règle, il est énuméré et devient donc candidat pour ce groupe de notes. On note que l'architecture de règles proposée est ouverte, dans le sens où elle laisse place à la reconnaissance d'autres types d'accords, pour peu que l'on définisse les règles associées. Aucun conflit entre les règles n'est à envisager, puisque un accord qui est éligible selon une règle sera bien énuméré, même s'il n'est pas éligible selon une autre. Nous présentons ci-après les règles que nous proposons, en fonction des types d'accords qu'elles reconnaissent. Leur intégration progressive dans le système fera ensuite l'objet d'une évaluation.

- triades majeures/mineures : un accord de ce type est candidat si chaque note du groupe de notes appartient à la triade considérée. Par exemple, le groupe de notes (do,mi) accepte deux accords candidats : La mineur et Do Majeur.
- accords de 7ème (majeurs/mineurs) : un accord de ce type est candidat si chaque note du groupe de notes appartient à l'accord et si la fondamentale et la 7ème de l'accord sont présentes dans le groupe de notes. Cette dernière règle permet d'éviter d'avoir l'accord de Do Majeur^7 candidat pour les groupes de notes tels que (Mi,Sol,Si), ou (Sol,Si) qui indiquent plutôt des accords de Mi mineur, ou Sol Majeur.
- accords de 9ème (majeurs/mineurs) : un accord de ce type est candidat si chaque note du groupe de notes appartient à l'accord et si la fondamentale, la quinte et la 9ème de l'accord sont présentes dans le groupe de notes.
- accords de 11ème (majeurs/mineurs) : un accord de ce type est candidat si chaque note du groupe de notes appartient à l'accord et si la fondamentale, la quinte et la 11ème de l'accord sont présentes dans le groupe de notes.
- d'autres types d'accords peuvent être recherchés, comme le type *min75b*. Pour que de tels types soient reconnus, toutes les notes composant l'accord doivent être présentes dans le groupe de notes (par exemple, un accord de type *min75b* est reconnu, et devient ensuite candidat, si la fondamentale, la tierce mineure, la quinte diminuée et la septième mineure sont dans le groupe de notes étudié).

On notera que les types d'accords présentés dans la liste de règles ci-dessus ne sont utilisés qu'à des fins de reconnaissance de motifs : il est ainsi possible de reconnaître, grâce à ces règles, un accord de Do Majeur^9, mais de n'énumérer qu'un accord en triade de Do Majeur, si l'on souhaite que la méthode ne produise qu'une séquence d'accords en triade (majeur/mineur), ce qui sera le cas dans les évaluations présentées par la suite, sauf mention contraire.

D'après ces règles de recherche de type d'accord, il est possible de rencontrer un groupe de notes pour lequel aucune tonalité et/ou aucun accord n'est énumérable. Dans ce cas, nous proposons le fonctionnement suivant :

- si aucun accord n'est énuméré pour le n-ième groupe de notes de la pièce, les candidats harmoniques énumérés sont les accords candidats du $(n-1)$-ième groupe de notes, combinés avec les tonalités énumérées pour le n-ième groupe de notes de la pièce, ceci pour permettre un possible changement de tonalité.
- si aucun accord et aucune tonalité ne sont énumérés pour le n-ième groupe de notes de la pièce, les candidats harmoniques énumérés pour ce groupe de notes sont les candidats du $(n-1)$-ième groupe de notes de la pièce.

5.2.3 Évaluation et résultats

Différentes évaluations du système appliqué à la musique symbolique sont présentées dans cette section. Suivant les évaluations, différentes bases de données ont été utilisées, ainsi que différents protocoles d'évaluation, qui sont présentés ci-après. Néanmoins, il convient de souligner qu'aucune évaluation n'a été possible pour l'estimation de la tonalité locale, ceci en raison du manque d'une base de données conséquente et annotée selon cette information tonale. Les évaluations du système pour une représentation symbolique de la musique concernent donc l'estimation des accords.

5.2.3.1 Distances musicales

Comme nous l'avons vu précédemment, différentes distances musicales, présentées en 3.1.3, peuvent être adoptées pour pondérer chaque arc du graphe de couples candidats, et ainsi influencer la sélection de la meilleure séquence de couples candidats pour la pièce étudiée. D'une distance musicale à une autre, le chemin optimal dans le graphe peut ainsi être différent, engendrant une séquence d'accords et de tonalités locales estimée différente.

Nous nous proposons d'évaluer l'efficacité de chacune des distances musicales présentées en 3.1.3. Pour cela, nous considérons successivement les distances musicales précédemment exposées comme pondération des arcs du graphe de candidats.

Base de données

57 morceaux d'accompagnement, générés par le logiciel Band-In-A-Box [1] et au format MIDI sont utilisés pour cette évaluation (voir A.1.3). Le nombre restreint de morceaux s'explique par le fait que nous cherchons à encourager les différences potentielles induites par les différentes distances musicales étudiées. Nous sélectionnons donc des pièces pour lesquelles le nombre moyen de couples candidats par groupe de notes était important, de

1. `http://www.pgmusic.com/products\_bb.htm`

manière à créer un maximum de diversité dans le graphe de candidat, pour ainsi engendrer une très grande quantité de chemins différents et donc encourager différentes séquences d'accords en fonction des différentes distances musicales étudiées.

Procédure d'évaluation

L'évaluation des estimations d'accords se fait en utilisant le *tick* MIDI, qui est l'unité de temps du format MIDI. Pour chaque *tick* MIDI, l'accord proposé par le système est comparé avec l'accord issu des annotations du logiciel Band In A Box. Ce processus est répété pour chaque morceau de la base de données, et pour chaque distance musicale étudiée. Trois types de scores différent sont distingués :

- un score de *fondamentale* : pour chaque *tick* MIDI, ce score vaut 1 si l'accord estimé possède la même fondamentale que l'accord annoté, et 0 dans les autres cas,
- un score de *fondamentale et mode* : pour chaque *tick* MIDI, ce score vaut 1 si l'accord estimé possède la même fondamentale et le même mode que l'accord annoté, et 0 dans les autres cas,
- un score de *fondamentale, mode et type* : pour chaque *tick* MIDI, ce score vaut 1 si l'accord estimé possède la même fondamentale, le même mode et le même type que l'accord annoté, et 0 dans les autres cas. Pour le calcul de ce score, seuls deux types différents sont considérés : le type *triade* (fondamentale, tierce et quinte), et le type *non-triade* (accords de 7ème, de 9ème, ...).

Résultats

Le tableau 15 présente les scores des différentes distances musicales intégrées dans le système proposé. Ces résultats ont également fait l'objet d'une étude de similarité, en utilisant la méthode statistique de Tukey-Kramer (*HSD multi-comparison*), comme proposé par le MIREX[2] sur les scores morceau par morceau, afin de constater si les distances étudiées donnent des résultats significativement différents les uns des autres.
Une représentation graphique de ces similarités est présentée sur la figure 60, et les résultats du test de différence significative est présenté sur le tableau 16.

Une première analyse des scores des différentes distances musicales nous permet de constater que les distances de Lerdahl (asymétrique et symétrique) se révèlent être les plus performantes pour le système proposé, suivies de la distance de Paiement. Néanmoins, lorsque l'évaluation porte non plus seulement sur la fondamentale et le mode, mais également sur le type des accords estimés, ces distances ne sont plus les plus performantes. À l'inverse, les distances d'Estrada et de contenu d'intervalle, appartenant toutes deux

2. `http://www.music-ir.org/mirex/2008/index.php/MIREX2008\_Results`

	Fondamentale+ mode+type	Fondamentale+ mode	Fondamentale seule
Lerdahl	69.03	**85.97**	**87.05**
Spirale	69.29	83.27	85.33
Estrada	**74.75**	78.66	83.08
Contenu d'intervalle	71.21	78.55	82.62
Paiement	68.67	85.29	86.7
Costère	61	69.64	72.6
$\text{Lerdahl}_{symtrique}$	69.08	**85.97**	**87.05**
Pitch-Interval	65.95	76.61	84.2
Cycle des quintes	72.38	81.28	86.42

Table 15 – Résultats de l'estimation d'accords selon la distance musicale utilisée comme calcul de coût dans le système proposé.

aux 3 moins bonnes distances pour la fondamentale seule et la fondamentale combinée au mode, se montrent cette fois les plus performantes lorsque le type d'accord est considéré dans l'évaluation. Il est également intéressant de noter que la distance issue du cycle des quintes fait toujours partie des cinq meilleurs distances, quel que soit le score considéré, en se hissant même au second rang sur l'évaluation combinant fondamentale, mode et type.

Si l'on cherche maintenant à analyser les ressemblances entre ces distances musicales, il apparaît un certain nombre de similitudes. Au delà de la ressemblance (très prévisible) entre les deux distances de Lerdahl (symétrique et asymétrique), ces deux distances se comportent de la même manière que la distance en spirale de Chew, et également, même si de manière moins évidente, que la distance de Paiement. Cette impression est confirmée par l'étude de similarité, qui met en avant le fait que ces 4 distances semblent appartenir à un même groupe. La distance de contenu d'intervalle, par définition proche de celle d'Estrada car fondée sur des considérations communes sur les intervalles, lui semble également statistiquement liée.

À la lumière de ces analyses, il nous apparaît que l'adoption de la distance de Lerdahl constitue le meilleur choix pour notre système, car en plus d'être la seule distance musicale conférant une influence à la tonalité locale, elle permet d'obtenir les meilleurs résultats pour l'estimation de fondamentale et de mode, qui constitue l'objectif principal, en terme d'évaluation, des travaux présentés ici.

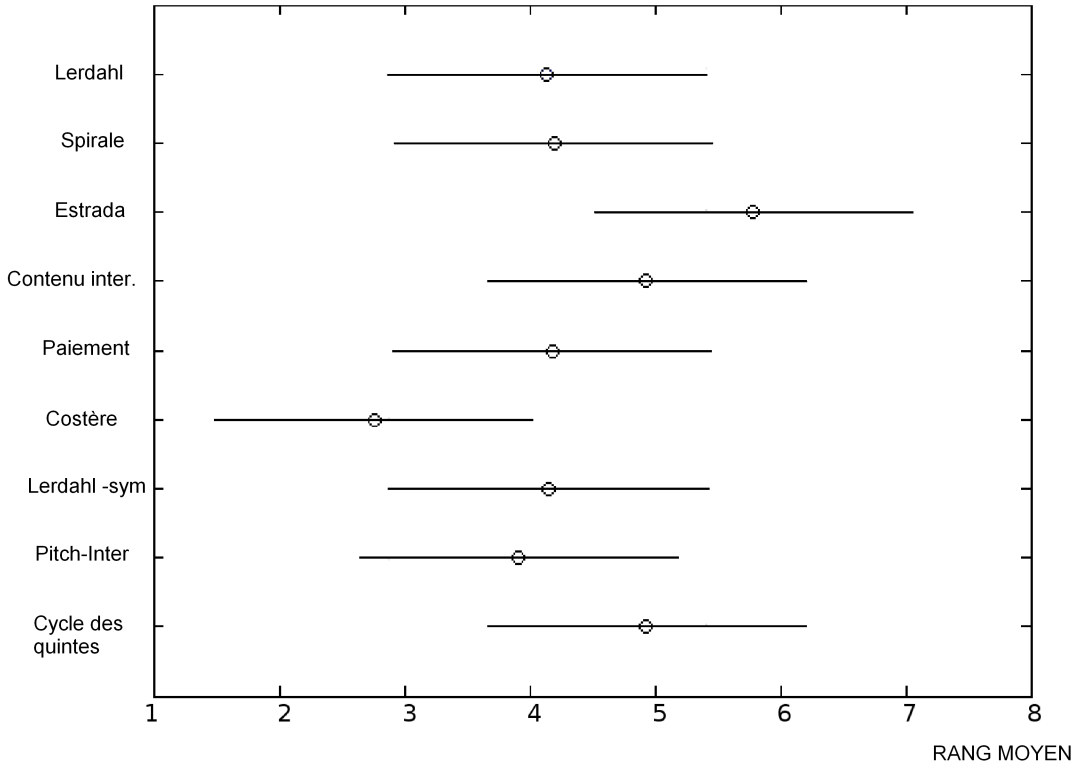


FIGURE 60 – Estimation de similarité des distances musicales selon le rang moyen (en abscisse), comme proposé par le MIREX.

	Lerd.	Spir.	Estr.	Int.	Paiem.	Cost.	L.sym.	P.-Int.	Cycle
Lerdahl	0	0	1	0	0	1	0	0	0
Spirale		0	1	0	0	1	0	0	1
Estrada			0	0	1	1	1	1	0
Contenu Interv.				0	0	1	0	1	0
Paiement					0	1	1	0	0
Costère						0	1	1	1
Lerdahl_{sym}							0	0	0
Pitch-Interval								0	1
Cycle									0

TABLE 16 – Test de différence statistique significative des distances musicales.

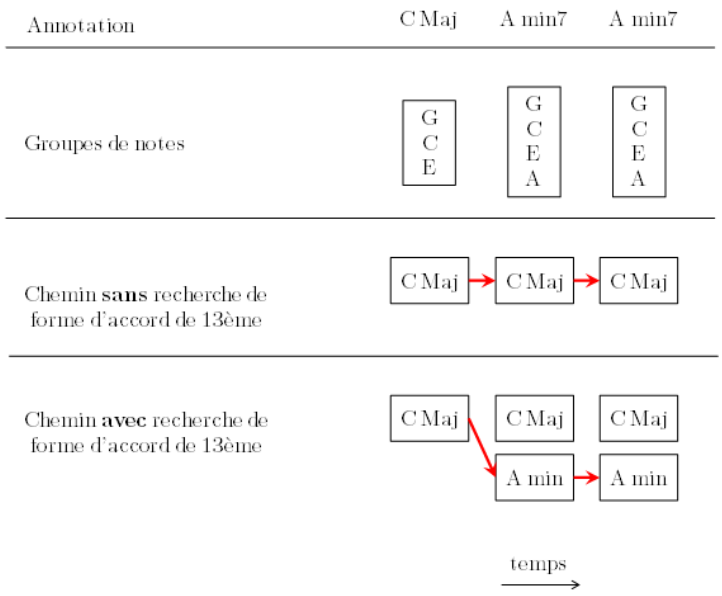


Figure 61 – Exemple d'extension de règle d'énumération causant une erreur dans l'estimation d'accords. Le chemin optimal qui correspond le mieux aux annotations est ainsi obtenu en excluant la recherche de type d'accord de 13ème.

5.2.3.2 Évaluation de l'énumération des candidats

Pour évaluer l'influence des règles d'énumération des accords et des tonalités candidates, nous nous proposons de tester successivement l'efficacité du système proposé, en étendant progressivement l'application des règles. L'ajout progressif des règles permettra en effet à un nombre de plus en plus important d'accords d'être énumérés pour un groupe de notes donné, mais une balance doit être trouvée, car l'énumération excessive d'accords occasionne des erreurs : énumérer un nombre important d'accord augmente en effet la probabilité de conserver le même accord énuméré à travers un changement d'accord, ce qui incite le système à continuer à choisir cet accord dans le chemin optimal, comme illustré sur la figure 61. Un compromis doit ainsi être défini en étendant le nombre de règles permettant l'énumération d'accords candidats fondée sur les recherches de motifs. Pour définir ce compromis, nous présentons l'expérience suivante.

Base de données

148 morceaux d'accompagnement, générés par le logiciel Band-In-A-Box au format MIDI sont utilisés pour cette évaluation, et sont regroupés en 4 styles musicaux distincts (voir annexe A.1.2).

Procédure d'évaluation

L'évaluation de l'estimation d'accords se fait morceaux par morceaux, et *tick* MIDI par *tick* MIDI : on compare, pour chaque *tick* MIDI, la fondamentale de l'accord proposé par le système avec la fondamentale de l'accord issu des annotations du logiciel Band In A Box. Ce processus est répété pour chaque *tick* d'un morceau donné, pour obtenir le score de la méthode évaluée sur ce morceau. On distingue ensuite deux manières de cumuler les résultats des différents morceaux au sein de la base de données

- une moyenne par *tick* (M.T.) qui pondère chaque morceau par son nombre de *tick*s. Les morceaux plus longs ont ainsi plus de poids que les morceaux courts.
- une moyenne par morceau (M.M), qui confère le même poids à chaque morceau.

Ces deux types de score sont calculés sur chaque style musical, et ce pour chaque configuration de règles introduite dans le système. Voici les configuration évaluées (on se rapportera à la description de ces règles section 5.2.2.2) :

- r5 : seule la règle de recherche des types d'accords en triades est intégrée,
- r7 : on ajoute à la configuration précédente la règle de recherche des types d'accords de 7ème,
- r9 : on ajoute à la configuration précédente la règle de recherche des types d'accords de 9ème,
- r11 : on ajoute à la configuration précédente la règle de recherche des types d'accords de 11ème,
- r13 : on ajoute à la configuration précédente la règle de recherche des types d'accords de 13ème.

Résultats

On trouvera les résultats tableau 17. L'efficacité maximale du système semble être atteinte pour la configuration r11. On notera que le passage de la configuration r5 à la configuration r7 présente une amélioration importante de l'efficacité générale de l'estimation. Les styles jazz et latins, semblent profiter le plus de cette amélioration (les scores de ces deux styles sont augmentés respectivement de 16% et 19%). Cela s'explique principalement par la présence importante d'accords de 7ème dans ces deux styles musicaux, alors que la reconnaissance de type d'accords de 7ème constitue la différence entre les configurations r5 et r7.

La prise en compte de règles de recherche des types d'accords de 9ème (configuration r9) semble uniquement profitable à ces deux mêmes styles, leurs scores augmentant encore de 4%. La configuration r11, ne propose, elle, qu'une faible amélioration de l'efficacité générale.

	r5		r7		r9		r11		r13	
	M.T.	M.M.	M.T.	M.M.	M.T.	M.M.	M.T.	M.M.	M.T.	M.M.
latin	60.6	66.4	79.4	82.4	83.6	86.8	**84.1**	**87.4**	76.4	82.5
classique	71.7	72.4	86.5	86.0	**87.0**	**86.4**	86.8	86.3	85.0	85.6
pop	79.7	76.2	86.1	85.2	87.2	86.8	**87.4**	**87.2**	85.4	84.0
jazz	53.4	52.4	69.0	68.2	72.9	72.0	**73.1**	**72.3**	67.9	66.0
Total	63.3	63.7	77.6	77.4	80.4	80.6	**80.7**	**80.9**	76.2	76.2

Table 17 – Pourcentage d'efficacité des différentes configurations de règles pour l'estimation des fondamentales des accords. La meilleure configuration est la configuration r11.

Enfin, la configuration r13 voit une chute notable de la précision générale, en particulier pour les styles jazz et latin (avec des chutes respectives de 5% et 8%). Une explication possible peut être apportée par un simple exemple. Considérons le groupe de notes (La,Do,Mi,Sol). Pour la configuration r11, seul l'accord de La Mineur est énuméré. Mais pour la configuration r13, l'accord de Do Maj 13 (sans 7ème, 9ème ou 11ème) est également énuméré. Les erreurs découlant de cet exemple simple peuvent ensuite être illustrées par la figure 61. Pour cette raison, la configuration conservée par la suite est la configuration r11.

5.2.3.3 Comparaison à *Melisma*

Le programme *Melisma Music Analyzer*, proposé par Temperley et Sleator [TS] se propose d'estimer la fondamentale des accords d'une pièce de musique au format MIDI. Ce programme implémente la méthode proposée par Temperley pour l'estimation de la fondamentale des accords, décrite précédemment. Une première phase d'estimation de métrique rythmique est ainsi menée, avec que chaque segment musical ne soit défini et annoté par une fondamental en utilisant un ensemble de règles dérivées des travaux de Lerdahl. Temperley a estimé la précision de ce programme sur les 48 fugues extraites du *Clavier bien tempéré* par J.S. Bach. L'efficacité annoncée de l'estimation des fondamentales sur cette base de données est de 83.7% [Tem99a].

Puisque *Melisma* ne fournit que la fondamentale des accords estimés, nous ne comparons dans un premier temps que l'estimation des fondamentales du système proposé avec l'estimation des fondamentales proposée par *Melisma*, selon le protocole d'évaluation décrit en 5.2.3.2. Dans un second temps, nous présentons une évaluation du système

	Système proposé		Melisma	
	M.T.	M.M.	M.T.	M.M.
latin	**87.0**	**89.1**	85.2	86.1
classique	**86.1**	**85.5**	80.7	80.3
pop	**89.1**	**89.5**	88.5	88.8
jazz	76.9	**77.3**	**77.6**	75.7
Total	**83.4**	**84.1**	82.9	82.2

TABLE 18 – Évaluation de l'estimation des fondamentales des accords. Le système proposé présente une efficacité légèrement supérieure à *Melisma*.

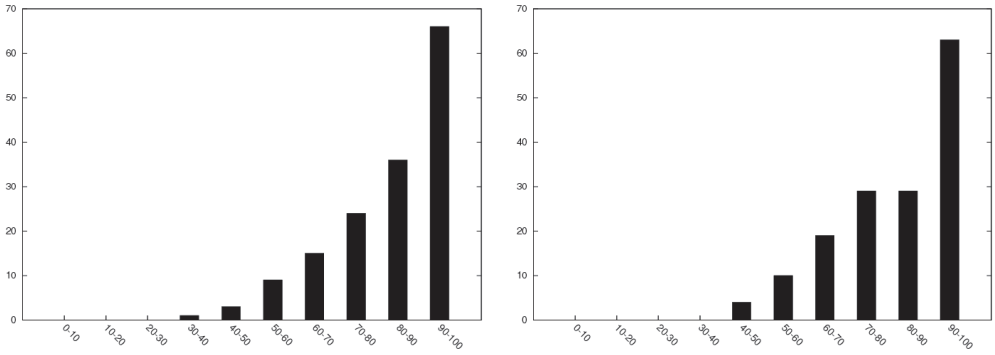


FIGURE 62 – Histogrammes représentant le nombre de morceaux (en ordonnée) par intervalle d'efficacité sur l'estimation des accord (en %, en abscisse), pour le système proposé (à gauche) et *Melisma* (à droite). Les deux systèmes présentent des profils différents.

proposé pour les fondamentales et les modes des accords estimés. Dans cette seconde évaluation, pour chaque *tick* MIDI, l'accord estimé doit présenter la même fondamentale et le même mode que l'accord annoté. Si l'accord annoté n'a pas de mode défini (et est, par exemple, de type suspendu, où la tierce est remplacé par la neuvième ou la quarte), seules les fondamentales sont comparées.

La comparaison de précision entre le système proposé et *Melisma* est présentée table 18. Si *Melisma* se montre plus précis sur le style *jazz*, (et uniquement selon la moyenne par *tick*), le système proposé se montre plus précis sur l'ensemble des autres styles, que ce soit selon la moyenne *tick* ou la moyenne morceau. La différence est particulièrement notable sur les styles *classique* et *latin*. Sur la moyenne globale des résultats, la différence est faible (moins d'1% sur la moyenne *tick* globale, et moins de 2% sur la moyenne morceau globale), mais toujours en faveur du système proposé.

	Système proposé	
	M.T.	M.M.
latin	86.8	88.9
classique	84.2	83.2
pop	88.8	**89.1**
jazz	75.3	76.0
Total	82.5	83.3

Table 19 – Évaluation de l'estimation des fondamentales et des modes des accords.

Ces résultats peuvent également faire l'objet d'une étude statistique. La figure 62 présente la répartition des morceaux selon le score de l'estimation d'accords pour les deux systèmes comparés. En abscisse se trouvent les différents intervalles de score, et en ordonnée se trouve le nombre de morceaux dont les scores se situent dans ces intervalles. On observe que les deux profils, bien que présentant certaines similitudes, sont loin d'être identiques. Le score le plus bas est ainsi de 41.9% pour *Melisma*, alors qu'il est de 36.5% pour le système proposé. Le score le plus élevé est de 100% pour les deux systèmes, mais pour des morceaux différents. Sur les 148 morceaux, le système proposé en analyse 102 avec une efficacité supérieure ou égale à 80%, alors que *Melisma* n'en présente que 92. Plus remarquable, les trois morceaux présentant la plus grande différence de score entre les deux systèmes sont tous trois mieux analysés par le système proposé : *April Joy* (90% contre 51.2%), *A Taste Of Honey* (86.4% contre 42.1%) et *The Entertainer* (95.5% contre 41.2%). On notera que ces trois morceaux appartiennent à trois styles musicaux différents. *Melisma* présente de son coté également certaines différences de précision en sa faveur, notamment sur *Alone Together* (84.2% contre 63.7%) et *April Showers* (95.3% contre 75.6%).
De telles statistiques montrent que même si les deux méthodes obtiennent des scores comparables, elles semblent utiliser différents types d'informations musicales. L'utilisation d'une analyse de la métrique et la gestion des notes de passages et ornements, qui constituent les différences majeures entre ces deux systèmes, peut expliquer une telle différence.

Le tableau 19 présente les résultats de l'évaluation du système sur la fondamentale et le mode des accords estimés. Sur les styles *pop, latin* et *classique*, les résultats sont légèrement meilleurs que les résultats de *Melisma* simplement évalué sur la fondamentale. Le résultat global calculé avec une moyenne sur les morceaux est également légèrement meilleur que ce même résultat pour *Melisma* simplement évalué sur la fondamentale. On peut ainsi affirmer que le système proposé se montre aussi précis que *Melisma*, mais en proposant en plus de la fondamentale, le mode des accords estimés.

5.2.4 Conclusion

Le système d'estimation d'accords depuis une représentation symbolique que nous proposons présente une précision comparable à celle de l'état de l'art, et propose non seulement la fondamentale, mais ajoute l'information essentielle sur le mode des accords estimés, ainsi qu'une estimation des tonalités locales.
Ces résultats sont prometteurs, notamment dans la perspective d'étendre les accords supportés par ce système, et d'utiliser, comme *Melisma*, des règles d'analyse de notes de passages et d'ornements. L'estimation des tonalités locales devra également faire l'objet d'une évaluation, mais une base de données comportant ces informations devra être constituée, car aucune n'existe pour l'heure à notre connaissance.

La proposition d'un système d'estimation simultanée d'accords et de tonalités locales depuis une représentation symbolique de la musique a fait l'objet d'une publication en conférence internationale [RRHS09], tout comme l'étude des différentes distances musicales au sein de ce système [RRHDC10].

5.3 Application depuis le signal audio

Après avoir étudié l'application à la musique symbolique, nous présentons dans cette section l'application de notre méthode dans le cas où l'entrée musicale analysée est un signal audio. Après avoir décrit le choix pour le calcul des vecteurs chroma, nous étudierons l'influence des différents paramètres du système (taille de la fenêtre d'analyse, filtrage, origine et nombre des candidats). Une évaluation de la méthode et la comparaison avec une méthode de l'état de l'art sera également proposée, avant que nous ne mettions en évidence l'influence bénéfique de l'estimation simultanée des accords et des tonalités locales.

5.3.1 Calcul de vecteurs chroma

Les *HPCPs* (*Harmonic Pitch Class Profiles*) désignent un type de vecteurs chroma proposé en 2006 par Gòmez [Góm06], et adopté par la suite notamment pour la recherche de reprises [SGHS08]. Les *HPCPs* sont inspirés des *PCPs* (*Pitch Class Profiles*) proposés en 1999 par Fujishima [Fuj99], et contiennent une prise en compte des harmoniques des fréquences détectées dans le signal analysé (voir figure 63). Le processus d'obtention d'un chromagramme de *HPCPs* est illustré figure 64. Nous utiliserons dans la suite une implémentation des *HPCPs* réalisée par nos soins.

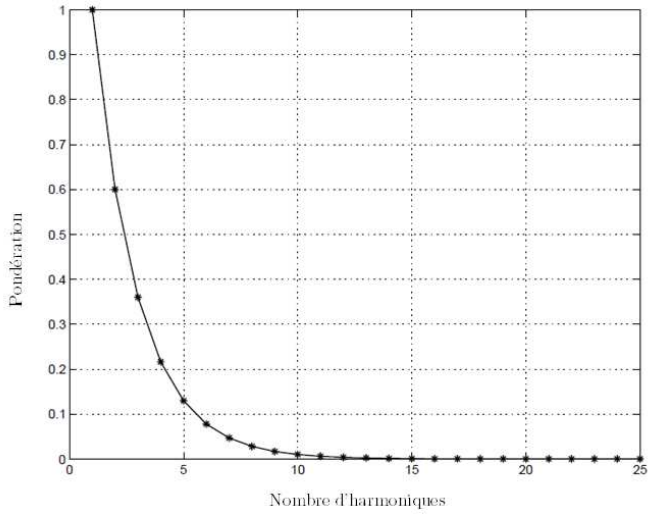


Figure 63 – Pondération des fréquences harmoniques pour le calcul des *HPCPs*, d'après [Góm06].

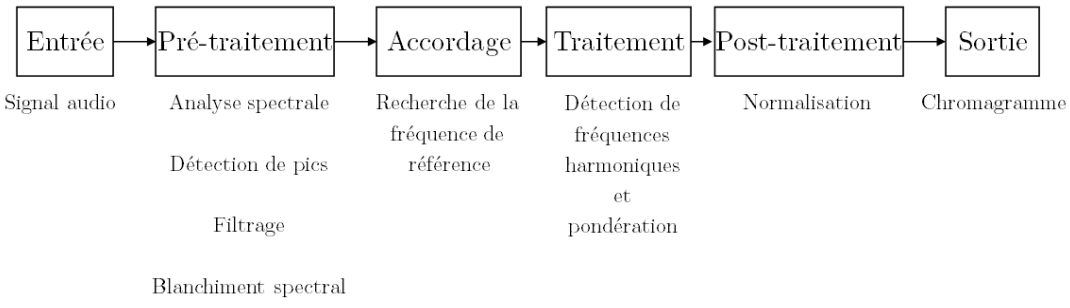


Figure 64 – Processus d'obtention d'un chromagramme de *HPCPs*, d'après [SGHS08].

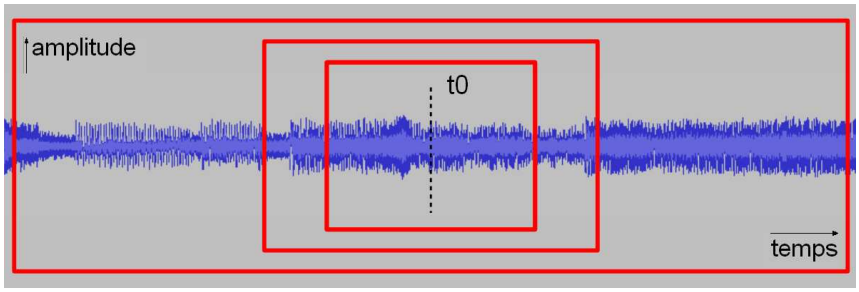


FIGURE 65 – Illustration de l'approche multi-résolutions : pour le même instant t_0 symbolisé par le segment vertical en pointillés, plusieurs tailles de trames peuvent être utilisées, définissant ainsi plusieurs résolution différentes.

5.3.2 Énumérations des candidats

Dans ce système, les accords et les tonalités sont énumérés de la même manière. Une corrélation est calculée, sous forme d'un produit scalaire, entre un vecteur chroma et des vecteurs modèles, représentant selon les cas des profils d'accords ou de tonalités (voir section 5.3.2.2). L'énumération des candidats, que ce soit pour les accords ou pour les tonalités locales, se fait ensuite en choisissant ceux ayant obtenu les scores de corrélation les plus importants avec les vecteurs chroma. Dans l'approche multi-résolutions présentée en 5.3.2.1, il est possible de considérer différentes trames centrées au même instant pour combiner les accords les plus fortement corrélés sur les différents vecteurs chroma correspondant. Ces différents accords candidats peuvent être porteurs d'informations différentes, ce qui pourra profiter au système. Plusieurs configurations, utilisant un nombre différents de candidats, et une origine différente des vecteurs chroma considérés sont décrites et évaluées dans la section 5.3.3.2.

5.3.2.1 Vecteurs chroma et approche multi-résolutions

Contrairement aux travaux existants dans ce domaine, qui n'utilisent qu'une taille unique de fenêtre d'analyse pour l'estimation des accords, nous proposons ici une approche combinant plusieurs résolutions, et utilisant donc plusieurs chromagrammes différents (et donc plusieurs tailles de trames différentes) pour le même signal audio (voir figure 65). L'intérêt principal d'une telle approche réside dans le fait que différentes résolutions peuvent apporter des informations différentes et complémentaires. Les vecteurs chroma calculés sur des trames plus longues pourront ainsi être plus utiles quand à l'estimation de tonalité, alors que des trames plus courtes seront plus utiles pour percevoir différentes variations tonales selon, par exemple, le tempo de la pièce de musique analysée.

5.3.2.2 Vecteurs modèles d'accords et de tonalités

Nombre de travaux [HS05, Lee06, OGF09] ont déjà utilisé des vecteurs modèles d'accords en triades, afin de déterminer la probabilité de leurs occurrences pour un chroma donné. En dimension 12, ces vecteurs modèles d'accords peuvent être définis de la manière suivante :

$$\text{Triade Majeure} = (1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0)$$

$$\text{Triade mineure} = (1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0)$$

Les 12 vecteurs modèles d'accords majeurs peuvent être obtenus à partir du modèle majeur par rotation d'indice. La même méthode permet d'obtenir les 12 vecteurs modèles d'accords mineurs à partir du modèle mineur.

Pour les tonalités locales, les vecteurs modèles utilisés sont les profils de tonalités proposés par Temperley [Tem99a] :

$$\text{Tonalité Majeure} = (5, 2, 3.5, 2, 4.5, 4, 2, 4.5, 2, 3.5, 1.5, 4)$$

$$\text{Tonalité mineure} = (5, 2, 3.5, 4.5, 2, 4, 2, 4.5, 3.5, 2, 1.5, 4)$$

Ces vecteurs modèles sont de dimension 12, et sont donc adaptés à une comparaison avec des vecteurs chroma de dimension 12. Si les vecteurs chroma sont de dimension 36, on se propose de transformer les vecteurs modèles ci-dessus en dimension 36, en concaténant les triplets $(0.75 \times c_i, c_i, 0.85 \times c_i)$ où c_i est une composante du vecteur modèle en dimension 12. Ce triplet a été constitué empiriquement sur la base de donnée des Beatles. La dissymétrie observée peut être expliquée par la tendance plus fort au sous-accordage (la fréquence de référence est légèrement en dessous de 440 Hz pour le La e la troisième octave) qu'au sur-accordage.

Dans le cas où l'on considère des vecteurs de dimension 36, trois vecteurs modèles différents à une rotation d'indice près représentent un même accord. Parmi ces trois vecteurs, seul celui donnant la plus forte corrélation avec le vecteur chroma étudié est conservé pour établir le classement des accords et tonalités locales considérés en vue de l'énumération des candidats.

5.3.3 Évaluation

Base de données

Tous les résultats présentés dans cette section proviennent d'une évaluation de précision des accords et des tonalités locales menée sur l'intégrale des enregistrements studio des Beatles, dont la liste complète est proposée annexe A.3, et dont la transcription des séquences d'accords et de tonalités est proposée par le *Centre for Digital Music (C4DM)* de l'université londonienne du *Queen Mary* et disponible sur le site *www.isophonics.net*.

Protocole d'évaluation

Le dictionnaire d'accords utilisé dans les évaluations est constitué par les 24 accords de triades majeures et mineures. Ce choix s'appuie notamment sur la fréquence d'apparition de ces accords dans la musique occidentale, et dans la musique pop en particulier : selon une récente étude [BWF11], les accords de triades représentent 63 % des accords utilisés pour annoter une base de données de plus d'un millier de titres issus des têtes de ventes aux États-Unis depuis plus de 50 ans (et regroupant ainsi plusieurs styles). Si on leur ajoute les accords de septième, ces accords représentent plus de 85 % des accords utilisés pour les transcriptions. Les accords de septième sont proches des accords de triades car trois des quatre notes qui les composent constituent une triade. À la lumière de ces chiffres, il est ainsi possible d'avancer que les accords de triades représentent un moyen efficace de représenter la musique populaire.

Les accords mentionnés dans les transcriptions des titres de la base données sont issus d'un dictionnaire étendu [HS05], qui dépasse les seuls accords de triades majeures/mineures. Les annotations ont donc été adaptées à l'évaluation de triades : si un accord présente une tierce majeure (respectivement mineure), il est considéré comme étant une triade majeure (respectivement mineure). S'il ne présente pas de tierce, seule la fondamentale de l'accord est évaluée. Les silences ont été ignorés. L'accord estimé est considéré comme correct s'il présente la même fondamentale et le même mode que l'accord mentionnée dans la vérité terrain.

L'évaluation des accords est réalisée toutes les 185 ms (soit tous les 8192 échantillons, pour un signal échantillonné à 44100Hz). L'accord estimé à chaque date d'évaluation est comparé à la vérité terrain mentionnée dans les annotations, pour la même date. Le score final pour chaque morceau est le rapport entre le nombre de dates d'évaluation pour lesquelles l'accord estimé est correct et le nombre de dates d'évaluation totales. La procédure d'évaluation pour les tonalités est identique à celle des accords. Le score sur la base de données entière est la moyenne du score de tous les morceaux.

En plus de l'évaluation de la précision des séquences d'accords et de tonalités estimées par le système, nous proposons également une mesure de l'efficacité de l'énumération des candidats, c'est-à-dire des hypothèses d'accords et de tonalités pour chaque trame, en introduisant le *potentiel énuméré*, qui correspond au rapport des trames pour lesquelles l'accord correct est présent dans les candidats énumérés sur le nombre de trames totales. Ce potentiel énuméré correspond au maximum de performance du système, puisque ce dernier ne peut pas choisir le candidat correct s'il n'est pas énuméré.

5.3.3.1 Paramètres des vecteurs chroma

Dimension

Un premier aspect important du calcul des vecteurs chroma est la robustesse à l'accordage. En effet, si la fréquence standard du La est de 440Hz (à la troisième octave), il est néanmoins possible de rencontrer des enregistrements pour lesquels cette fréquence de référence présente une valeur différente, et même dans certains cas, une valeur qui diffère au cours du temps dans un même morceau. Certains travaux [HS05, MD08] proposent l'adoption de vecteurs chroma de dimension 36 au lieu de 12, de manière à pouvoir prendre en compte la déviation provoquée par l'accordage non standard de certains enregistrements.

Une première évaluation, concernant la dimension des vecteurs chroma est ainsi menée : pour deux paramétrisations différentes, $p1$ et $p2$, nous nous proposons d'évaluer le potentiel énuméré d'un chromagramme, de taille de fenêtre 743 ms (avec un recouvrement de 75%) pour les accords, et de taille de fenêtre 10 sec pour les tonalités locales. Pour $p12$ comme pour $p36$, les candidats tonalités locales énumérés pour chaque trame sont les 3 tonalités les plus fortement corrélées avec le vecteur chroma de la tonalité locale. La seule différence entre $p12$ et $p36$ est donc la dimension des vecteurs chroma calculé : les vecteurs chroma sont de dimensions 12 pour $p12$, et de dimension 36 pour $p36$.
Nous considérons successivement comme accords candidats le premier ($a1$), les deux premiers ($a1 + a2$), les trois premiers ($a1 + a2 + a3$) et les quatre premiers accords ($a1 + a2 + a3 + a4$) selon le classement induit par la corrélation avec les vecteurs modèles d'accords choisis, pour les configurations $p12$ et $p36$, et observons le potentiel énuméré des configurations ainsi engendrées. Les résultats sont présentés tableau 20.

	$p12$	$p36$
$a1$	24.2	**58.3**
$a1 + a2$	34.8	**71.5**
$a1 + a2 + a3$	42.9	**78.6**
$a1 + a2 + a3 + a4$	49.6	**82.9**

Table 20 – Potentiel énuméré en fonction du nombre de candidats accords énumérés, pour deux dimensions de vecteurs chroma différentes.

On remarque que $p36$ se montre nettement meilleur que $p12$ en terme de potentiel énuméré. Le premier candidat ($a1$) énuméré est ainsi correct plus de deux fois plus souvent en utilisant des vecteurs chroma de dimension 36 qu'avec des vecteurs chroma de dimension 12 (58.3% contre 24.2%).

Encore plus notable, la configuration $p36$ avec un seul candidat ($a1$) est meilleure de près de 10% que $p1$ avec quatre candidats ($a1 + a2 + a3 + a4$), toujours en terme de potentiel énuméré (58.3% pour la combinaison $a1$ et $p36$, contre 49.6% pour $a1 + a2 + a3 + a4$ avec $p12$).

Au vu de ces résultats, le choix des vecteurs chroma de taille 36 semble donc s'imposer, et pour la suite, les vecteurs chroma seront toujours calculés en dimension 36.

Taille des trames

La taille de trames précédemment choisie (743 ms) correspond à une taille très fréquemment utilisée dans les méthodes de reconnaissances d'accords [OGF11, BP05, HS05, LS08, Lee06, MD08]. Nous nous proposons néanmoins de tester d'autres tailles de fenêtres pour estimer, toujours à l'aide du potentiel énuméré, l'influence de la taille des fenêtres.
De même que précédemment, nous plaçons successivement comme accords candidats le premier ($a1$), les deux premiers ($a1 + a2$), les trois premiers ($a1 + a2 + a3$) et les quatre premiers accords ($a1 + a2 + a3 + a4$) selon le classement induit par la corrélation avec les vecteurs modèles d'accords choisis, et ce pour les différentes tailles de trame étudiées. Les résultats sont présentés sur le tableau 21.

Taille des trames (en ms)	185	371	743	1486
$a1$	46.1	51.3	**58.3**	54.3
$a1 + a2$	56.1	60.6	**71.5**	66.5
$a1 + a2 + a3$	65.1	70.6	**78.6**	73.4
$a1 + a2 + a3 + a4$	67.9	75.4	**82.9**	77.2

TABLE 21 – Potentiel énuméré en fonction du nombre de candidats accords énumérés, pour différentes tailles de trames différentes.

Il apparaît clairement que les trames de taille 743 ms s'avèrent plus intéressantes, le potentiel énuméré étant plus important pour chacune des configurations différentes de candidats énumérés. Dans la suite, la taille des trames utilisée sera ainsi fixée à 743 ms, sauf mention contraire.

Filtrage

Les travaux menés avec Laurent Oudre [RRHO10, OGF11] ont mis en avant les effets bénéfiques du filtrage, et en particulier du filtrage médian sur le calcul des vecteurs chroma : un filtre médian agissant sur une période de quelques secondes permet de réduire l'influence

des fréquences parasites. Ce type de filtrage est couramment adopté dans la littérature [MD08, MND09, MD10, PP08]. Nous nous proposons de tester l'influence de ce filtrage sur le calcul des vecteurs chroma, en réalisant un filtre médian sur chaque composante des vecteurs chroma, en considérant les vecteurs chroma précédents et suivants dans le chromagramme. Le filtrage utilisé est d'ordre 8, ce qui, avec la taille de trame précédemment adoptée, correspond à 5.9 secondes. Le tableau 22 présente l'influence du filtre médian sur le potentiel énuméré.

filtrage	sans	médian (ordre 8)
$a1$	58.3	**68.4**
$a1 + a2$	71.5	**79.1**
$a1 + a2 + a3$	78.6	**85.5**
$a1 + a2 + a3 + a4$	82.9	**88.9**

Table 22 – Potentiel énuméré en fonction du nombre de candidats accords énumérés, avec et sans filtre médian.

Ces résultats montrent une augmentation du potentiel énuméré pour toutes les configurations de candidats. On notera en particulier une hausse de plus de 10% sur le meilleur candidat ($a1$). Cette augmentation est de plus supérieure à 6%, quelle que soit la configuration de candidats considérée.

De même que pour la dimension des vecteurs chroma et pour la taille de trame adoptées, il apparaît clair que le filtrage doit être utilisé dans le calcul des vecteurs chroma en vue de l'estimation des accords. Dans la suite, tous les vecteurs chroma feront ainsi l'objet d'un filtrage médian.

5.3.3.2 Nombre et origine des candidats

Après avoir présenté l'influence des différents paramètres sur le calcul de vecteurs chroma , et proposé une optimisation du potentiel énuméré, nous allons étudier l'efficacité du système en fonction du nombre et de l'origine des candidats accords. Le tableau 23 présente la précision du système en fonction du nombre d'accords candidats insérés dans le système pour chaque trame.

La première remarque, triviale, est que le score du système, par définition, égal au potentiel énuméré lorsqu'un unique candidat accord est énuméré. Une seconde remarque, également attendue, est que le potentiel énuméré augmente au fur et à mesure que les candidats sont ajoutés à l'énumération. Mais l'augmentation de ce potentiel énuméré se fait au prix de l'accroissement du nombre de candidats. La difficulté augmente donc également pour choisir le candidat correct dans l'énumération.

	potentiel énuméré	séquence d'accords estimée
$a1$	68.4	68.4
$a1 + a2$	79.1	**70.0**
$a1 + a2 + a3$	85.5	64.3
$a1 + a2 + a3 + a4$	88.9	59.3

TABLE 23 – Potentiel énuméré et précision du système, en fonction du nombre de candidats accords énumérés.

Dans le tableau 23, on observe ainsi que si le passage de 1 à 2 candidats fait augmenter le score du système, les passages successifs de 2 à 3 candidats, puis de 3 à 4 candidats, le font chuter. Cette chute s'explique en observant la nature de ces accords candidats. Le classement des accords selon la corrélation de leurs vecteurs modèles avec le vecteur chroma tend en effet à rassembler, à l'intérieur de ce classement, des accords dont les vecteurs modèles sont proches les uns des autres. Ainsi, des accords partageant la même fondamentale (comme Do Majeur et Do mineur), ou plusieurs notes en commun (comme Do Majeur et Mi mineur) seront, par définition, peu discriminés les uns par rapport aux autres par le calcul de la corrélation, et se retrouveront proches dans le classement des candidats potentiels. On a ainsi observé que, dans 80% des trames de la base de données, les deux premiers accords du classement ont une distance inférieure ou égale à 1 dans le cycle des quintes. Ce cycle des quintes étant un élément prépondérant du calcul de la distance musicale entre candidats (distance de Lerdahl utilisée dans notre système), cette proximité entre les accords candidats d'une même trame va tendre à créer une faible distinction entre eux, propice à engendrer des erreurs dans le calcul du chemin optimal.

Il semble donc que l'énumération d'accords candidats provenant d'une unique information musicale (à savoir la corrélation avec un vecteur chroma donné) ne permette pas d'étendre efficacement le potentiel énuméré en augmentant le nombre de candidats énumérés. C'est pour cette raison qu'une approche multi-résolutions, présentée en section 5.3.2.1 est par la suite expérimentée.
Pour ce faire, deux configurations, baptisées *bi-chroma* et *tri-chroma*, sont étudiées :

- configuration bi-chroma : l'accord le plus fortement corrélé avec le vecteur chroma de la trame de longueur 743 ms est énuméré. Une autre trame, de longueur 185 ms et centrée sur celle de 743 ms est utilisée pour générer un autre vecteur chroma. L'accord le plus fortement corrélé avec ce deuxième vecteur chroma est énuméré (s'il est différent du premier accord déjà énuméré).
- configuration tri-chroma : l'accord le plus fortement corrélé avec le vecteur chroma de la trame de longueur 743 ms est énuméré. Deux autres trames, de longueur 93 ms,

Configuration	Bi-chroma	Tri-chroma
Potentiel énuméré (%)	76.7	**79.1**
Nb d'accords candidats différents	1.23	1.36
Précision du système (%)	72.8	**73.7**

Table 24 – Potentiel énuméré, nombre d'accords candidats distincts énumérés, et précision du système pour les configurations bi-chroma et tri-chroma.

consécutives et centrées sur la trame de 743 ms sont utilisées pour générer deux autres vecteurs chroma. Les deux accords les plus fortement corrélés avec ces deuxième et troisième vecteurs chroma sont énumérés (s'ils sont différents du premier accord déjà énuméré).

Selon les trames, la configuration bi-chroma énumère 1 ou 2 accords différents, et la configuration tri-chroma entre 1 et 3 accords différents. Le but est d'appuyer le choix d'un accord dans le cas où différents vecteurs chroma associés à une même trame de 743 ms présentent une corrélation maximale avec lui, et de donner plusieurs candidats différents au système lorsque les informations apportées par ces vecteurs chroma sont différentes.

Ces deux configurations montrent une meilleure précision que les configurations n'utilisant qu'un unique chromagramme. Le nombre de candidats est plus également faible, dans les deux cas, que pour la configuration utilisant les 2 meilleurs candidats avec un chromagramme unique (respectivement 1,23 et 1,36 contre 2). La configuration tri-chroma est plus performante, que ce soit pour le potentiel énuméré que pour la précision du système, et sera adoptée par la suite.

5.3.3.3 Estimation des tonalités locales

L'énumération des tonalités locales candidates utilise des vecteurs chroma calculés sur des trames d'environ 10 secondes. Les trois tonalités dont les vecteurs modèles de tonalité sont les plus fortement corrélés avec un vecteur chroma donné sont énumérées pour la trame correspondante. Nous nous proposons de modifier cette approche, pour y incorporer une considération multi-résolutions. Trois configurations sont ainsi expérimentées :

- c_{10-20} : deux candidats tonalité sont énumérés (par maximum de corrélation) à partir d'un vecteur chroma calculé sur une trame de 10 secondes et deux autres (potentiellement différents) à partir d'un vecteur chroma calculé sur une trame de 20 secondes,
- c_{10-25} : deux candidats tonalité sont énumérés (par maximum de corrélation) à partir d'un vecteur chroma calculé sur une trame de 10 secondes et deux autres (potentiellement différents) à partir d'un vecteur chroma calculé sur une trame de 25 secondes,
- c_{10-30} : deux candidats tonalité sont énumérés (par maximum de corrélation) à partir

d'un vecteur chroma calculé sur une trame de 10 secondes et deux autres (potentiellement différents) à partir d'un vecteur chroma calculé sur une trame de 30 secondes.

Les résultats sont présentés tableau 25. On détaille les tonalités locales correctement estimées par le système, et on précise la nature de l'erreur pour les tonalités incorrectes, en les séparant en trois catégories : tonalités relatives (exemple : Sol Majeur est estimé à la place de Mi mineur), tonalités voisines (exemple : Fa Majeur est estimé à la place de Do Majeur), et les autres cas d'erreurs (lorsque la tonalité estimée n'est pas la tonalité correcte, ni la tonalité relative, ni une de ses tonalités voisines).

Tonalités estimées	Correctes	Relatives	Voisines	Autres
Sans multi-résolutions (%)	6.4	2.9	17.4	17.3
c_{10-20} (%)	**63.9**	2.9	16.4	**16.8**
c_{10-25} (%)	63.6	2.9	16.2	17.3
c_{10-30} (%)	63.1	2.9	16.3	17.7

TABLE 25 – Analyse de l'erreur sur l'estimation de tonalités locales.

On s'aperçoit ici que l'apport de l'approche multi-résolutions, tout en étant intéressant et fournissant une amélioration des résultats, n'est pas aussi bénéfique que pour les accords. Ceci peut s'expliquer par le nombre relativement restreint de changements de tonalité (en moyenne, moins de 2 par morceau) dans la base de données étudiée. Néanmoins, il convient de noter que l'approche multi-résolutions, et en particulier la configuration c_{10-20} permet non seulement d'augmenter la précision de l'estimation sur le score des tonalité correctes, mais également de minimiser la proportion d'erreurs lointaines, c'est-à-dire des erreurs pour lesquelles la tonalité estimée n'est ni correcte, ni relative, ni voisine.

5.3.3.4 Lissage final

La dernière étape de l'estimation de la séquence d'accords est un lissage, qui s'opère une fois que la meilleure séquence de candidats énumérés dans le graphe a été choisie par programmation dynamique. Sur cette séquence de candidats, certaines erreurs peuvent subsister pour des raisons musicales. Ainsi, une note de passage jouée lors un solo instrumental, ou sur une ligne mélodique chantée, peut occasionner un bruit harmonique sur une trame de signal isolée, qui peut provoquer un changement d'accord local. Sur la séquence d'accords finale, les motifs de la forme $AABAA$ sont remplacés par $AAAAA$. En appliquant un tel traitement, la précision du système atteint 75% d'efficacité pour l'estimation des accords sur la base de données considérée.

Méthode	Fondamentale	Fondamentale et Mode
OGF2 (%)	**78.9**	72.3
Système proposé (%)	77.9	**75.0**

Table 26 – Comparaison entre la méthode proposée et la méthode OFG2. Les scores affichés sont une moyenne de la précision par morceau sur la base de données.

5.3.3.5 Comparaison à une méthode de l'état de l'art pour les accords.

Nous comparons notre système à une des meilleures méthodes d'estimation d'accords ayant concouru au MIREX 2009, évaluée sur la même base de données et avec le même protocole d'évaluation. Il s'agit de la méthode OGF2, proposée par Oudre, Grenier et Févotte qui s'est classée première au MIREX 2009 pour la tâche *Audio Chord Detection : root*, et deuxième pour la tâche *Audio Chord Detection : root and mode*. On trouvera les résultats de cette comparaison dans le tableau 26. Pour l'estimation de fondamentale seule, OGF2 présente une précision plus haute d'environ 1% que la méthode proposée (78,9% comparé à 77,9%). Pour l'estimation de fondamentale et de mode, le système proposé se montre plus efficace que la méthode OGF2, et ce de près de 3% (75,0% comparé à 72,3%). Cette comparaison prouve que le système proposé offre une précision d'estimation d'accords comparable, et même supérieure à l'une des meilleurs méthodes proposées au MIREX 2009 pour l'estimation d'accords depuis le signal audio. La méthode proposée combine donc une estimation d'accord très intéressante, et offre simultanément une estimation des tonalités locales.

5.3.3.6 Comparaison à une méthode directe pour les tonalités locales

L'estimation de tonalités locales est menée sur la même base de données, et avec le même protocole d'évaluation. Nous comparons cette fois l'estimation de notre système à l'estimation d'une méthode directe, qui utilise les mêmes paramètres que la méthode proposée (taille de trame de 10 sec et mêmes vecteurs modèles de tonalité). Pour cette méthode directe, seule la tonalité la plus fortement corrélée sur une trame donnée est considérée, et donne directement la tonalité estimée sur cette trame. Les résultats, présentés dans le tableau 27, détaillent la relation entre la tonalité estimée et la tonalité de la vérité terrain.

Le système proposé présente une précision supérieure à la méthode directe, en estimant un plus grand pourcentage de tonalités correctes (63,9% comparé à 57,6%). De plus, le nombre d'erreurs dues à l'estimation d'une tonalité "lointaine" (non correcte, non relative et non voisine) est plus faible avec le système proposé qu'avec la méthode directe (16,8% comparé à 21,9%).

Tonalité estimée	Correcte	Relative	Voisine	Autre
Système proposé (%)	**63.9**	2.9	16.4	**16.8**
Méthode directe (%)	57.6	1.6	18.9	21.9

TABLE 27 – Le système proposé présente une précision supérieure à la méthode directe en termes de tonalités correctes détectées. Le nombre d'erreurs "lointaines" (autres) est également plus faible avec le système proposé.

Candidats	Accord		Tonalité		Accord et Tonalité	
	A	•	•	T	A	T
Système proposé (%)	73.1	•	•	57.8	**75.0**	**63.9**

TABLE 28 – Précision du système en considérant seulement les accords, seulement les tonalités, ou les deux simultanément comme candidats. Le système présente de meilleurs résultats lorsque les accords et les tonalités sont estimés simultanément.

5.3.3.7 Influence de l'estimation simultanée

Nous présentons ici une évaluation des bénéfices de l'estimation simultanée d'accords et de tonalités. Ces bénéfices sont mesurés en n'estimant successivement que les accords, puis que les tonalités, et enfin les deux simultanément. Pour cela, nous avons construit deux systèmes quasi identiques au système optimal présenté précédemment, mais présentant dans un cas seulement les accords comme candidats, et dans l'autre seulement les tonalités. La distance utilisée pour calculer la distance entre candidats est la distance issue du cycle des quintes. Les résultats sont présentés dans le tableau 28.

On note que l'estimation simultanée profite à la précision des accords comme à celle des tonalités. Sans cette estimation simultanée, la précision des accords chute en effet de près de 2% (75,0 comparé à 73,1) et l'estimation de tonalité de plus de 6% (63,9 comparé à 57,8).

5.3.4 Conclusion

L'adaptation du système proposé au signal audio présente une précision supérieure à une des méthodes de l'état de l'art les plus performantes pour l'estimation des accords, et une précision supérieure à une méthode d'estimation directe pour les tonalités locales.
Nous avons montré l'influence des différents paramètres du système, du calcul des vecteurs chroma à l'origine des candidats, en mettant notamment en évidence les apports d'une approche multi-résolutions, et le bénéfice réciproque de l'estimation simultanée des accords et des tonalités locales.

Une première version de cette méthode d'estimation conjointe d'accords et de tonalités depuis le signal audio a fait l'objet d'une soumission au MIREX 2009 pour la tâche *Audio Chord Detection*. La méthode soumise a terminé troisième en précision pour la détection de fondamentale et cinquième pour la détection de fondamentale et mode (sur huit participants). Cette méthode a également fait l'objet de publications en conférences internationales et nationales [RRHO10, Roc10].

5.4 Conclusion

Nous avons présenté dans ce chapitre une nouvelle méthode d'estimation d'accords et de tonalités locales. Cette méthode, appliquée à la musique symbolique, se montre plus précise que le logiciel *Melisma* pour l'estimation d'accords. Appliquée au signal audio, elle se montre plus précise qu'une des méthodes de l'état de l'art les plus performantes pour l'estimation d'accords, tout en estimant les tonalités locales.
Il semble maintenant intéressant de nous pencher sur les possibles applications de l'estimation des informations tonales, et en particulier sur l'étude de similarité musicale qui sera abordée dans le prochain chapitre.

Chapitre 6

Estimation de similarité harmonique

Ce chapitre présente une application des travaux présentés dans les chapitres précédents : la recherche de morceaux par similarité dans une base de données. Le cas de figure étudié est le suivant : un morceau appelé *requête* est comparé un à un à tous les autres morceaux de la base de données. Chaque comparaison entre cette requête r et un morceau m_i donne lieu au calcul d'un score de similarité. Plus ce score est élevé, plus les morceaux r et m_i sont estimés comme similaires. Tous les morceaux m_i sont ensuite classés selon leurs scores de similarité respectifs. Grâce à ce processus, il est possible d'utiliser des requêtes par fredonnement, où la requête est chantonnée par l'utilisateur, de créer des listes de morceaux homogènes en regroupant les titres présentant une forte similarité entre eux, ou encore de chercher à découvrir de nouveaux morceaux à partir d'un titre fourni par l'utilisateur.

Nous cherchons dans ce chapitre à utiliser les séquences d'accords pour estimer une similarité harmonique entre deux morceaux. Comme le montre la figure 66, la comparaison des séquences d'accords intervient après le processus d'estimation des accords. Nous nous proposons donc dans un premier temps d'étudier la similarité entre séquences d'accords annotées, avant d'intégrer le processus d'estimation des accords dans la recherche de similarité harmonique. Nous verrons ainsi dans la section 6.1 comment définir une mesure de similarité entre séquences d'accords, avant de présenter en section 6.2 l'étude de similarité harmonique utilisant les séquences d'accords estimées par le système présenté chapitre 5. Nous terminerons en montrant l'apport du modèle en arbre de l'harmonie présenté au chapitre 3. La prise en compte d'un tel modèle permet en effet une amélioration dans l'estimation de similarité, qui s'avère plus performante sur quelques cas de plagiat, comme nous le montrerons en section 6.3.

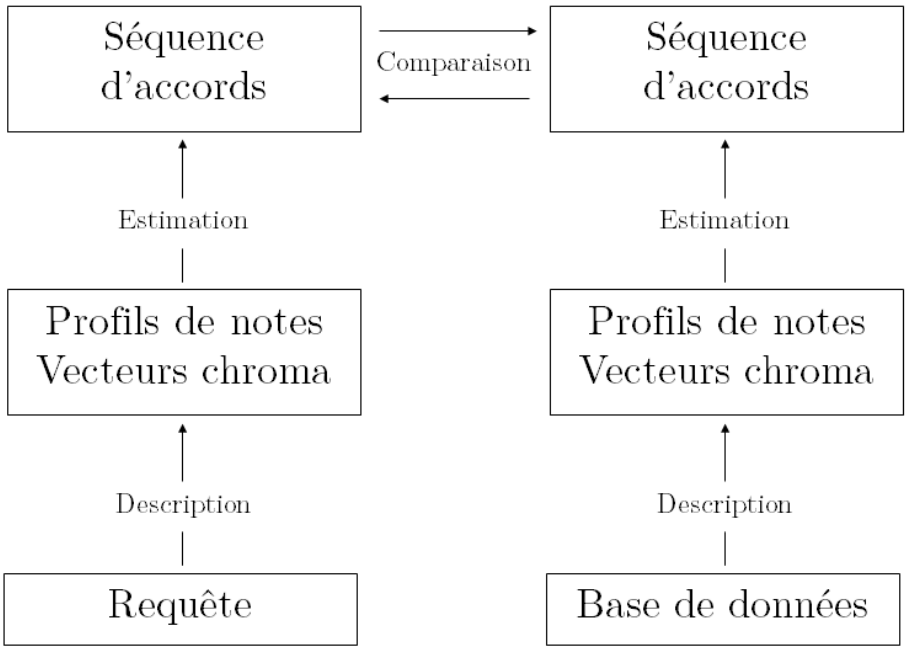


FIGURE 66 – Illustration de l'utilisation des séquences d'accords pour la recherche de similarité dans une base de données.

6.1 Similarité entre séquences d'accords

Cette section propose une étude de la comparaison de séquences d'accords. En pratique, ces dernières sont supposées connues, en utilisant les annotations des séquences d'accords dans les morceaux considérés. L'estimation de ces séquences d'accords n'est donc pas abordée pour l'instant (voir figure 66).

Si le nombre de méthodes proposées dans la littérature pour l'estimation d'accords est important, on dénombre en revanche très peu de travaux dans l'étude de similarité entre séquences d'accords. À notre connaissance, seuls les travaux de De Haas [DHVW08] s'inscrivent dans la comparaison entre deux séquences symboliques d'accords. La méthode proposée par De Haas, nommée *Tonal Pitch Step Distance*, se propose de représenter graphiquement l'aire formée, au cours du temps, par les distances musicales de Lerdahl entre les accords c_i d'une séquence d'accord c et l'accord correspondant au premier degré de la tonalité du morceau. La méthode superpose géométriquement les aires associées à deux séquences d'accords c et d, et cherche ensuite à minimiser la différence d'aire selon un décalage horizontal. On trouvera figure 67 une illustration de cette méthode.

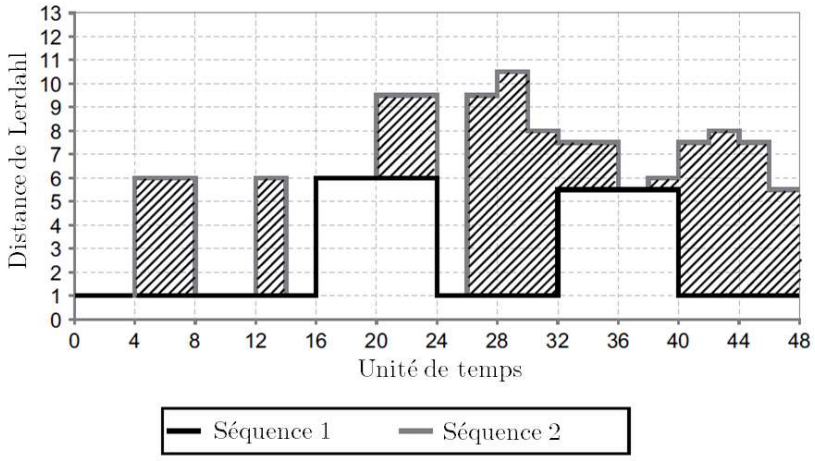


Figure 67 – Illustration de la méthode *Tonal Pitch Step Distance*, d'après [DHVW08]. Les aires correspondant aux deux séquences d'accords sont superposées, et la méthode cherche le minimum de différence d'aire.

6.1.1 Alignement local

Pour prétendre à une mesure de similarité musicale, nous devons proposer une méthode de comparaison prenant en compte certaines variations entre les séquences d'accords comparées. Dans l'exemple suivant, les deux séquences d'accords s1 et s2, toutes deux typiques du genre blues, présentent certaines variations (insertion de l'accord de G7 ou de C7, ou bien substitution d'un accord diminué par un accord de septième par exemple) tout en demeurant très proches harmoniquement parlant. Ces deux progressions présentent en effet toutes deux l'affirmation du premier degré lors de la première mesure, un passage du premier au quatrième degré à la cinquième mesure et le cinquième degré à la neuvième mesure.

```
s1:
GMaj  -  -  - | GMaj  -  -  - | GMaj  -  -  - | GMaj  -  -  -
CMaj  -  -  - | CMaj  -  -  - | GMaj  -  -  - | GMaj  -  -  -
DMaj  -  -  - | CMaj  -  -  - | GMaj  -  -  - | GMaj  -  -  -

s2:
GMaj  - G7  - | CMaj  - C7  - | GMaj  -  -  - |  G7   -  -  -
CMaj  -  -  - | Cdim  -  -  - | GMaj  -  -  - |  G7   -  -  -
 D7   -  -  - |  C7   -  -  - |  G7   -  -  - |  D7   -  -  -
```

Pour prendre en compte de telles variations, nous considérons des méthodes issues de l'algorithmique du texte, et en particulier les algorithmes d'alignement local, qui furent utilisés à des fins de comparaison mélodique [MS90, HFR07], de vecteurs chroma [SGHS08],

et également adaptés pour des comparaisons d'accords [Bel07]. Entre autres approches, la méthode proposée par Smith et Waterman [SW81] consiste à détecter des similarités locales entre deux séquences. Cette méthode localise et extrait des couples de régions similaires dans les séquences à comparer. Nous nous proposons d'adapter cette méthode pour la comparaison de séquences d'accords.

Pour calculer un score de similarité entre deux séquences u et v, on transforme u en v à l'aide d'opérations élémentaires, chacune d'elle étant associée à un score de similarité élémentaire [NW70]. L'ordre dans lequel interviennent ces opérations n'a pas d'influence. Le score de similarité total est la somme des scores induits par les opérations élémentaires. On distingue trois opérations élémentaires : la substitution, la suppression et l'insertion. Le score associé à l'insertion d'un accord est constant. Pour assurer la symétrie entre la transformation de aba en aa et la transformation de aa en aba par exemple, le score d'insertion est égal au score de suppression d'un accord. Ici, les scores d'insertion et de substitution sont égaux à -1. Nous nous proposons d'examiner plusieurs propositions pour définir le score $s(a_1, a_2)$ associé à la substitution d'un accord a_1 à un autre accord a_2 :

- un score binaire :
  $s(a_1, a_2) = 2$ si $a1$ et $a2$ présentent la même fondamentale
  $s(a_1, a_2) = -2$ sinon,
- un score tenant compte de la distance d_f entre les fondamentales de a_1 et a_2 sur le cycle des quintes :
  $s(a_1, a_2) = 2$ si $d_f = 0$
  $s(a_1, a_2) = -2$ si $d_f \geq 4$
  Sinon, $s(a_1, a_2)$ prend des valeurs décroissantes entre 2 et -2 suivant d_f (si $d_f = 1$, $s(a_1, a_2) = 1$, si $d_f = 2$, $s(a_1, a_2) = 0$, etc),
- un score dit de "consonance", tenant compte de la distance entre les fondamentales de a_1 et a_2, proposé par Hanna et al. [HFR07] et représenté tableau 29,
- un score inversement proportionnel à la distance de Lerdahl [Ler01], présentée en 3.1.3.

En plus des considérations sur la fondamentale des accords a_1 et a_2, nous nous proposons également de tenir compte des modes des accords considérés. Ainsi, si le mode est pris en compte, le score de substitution sera augmenté de 0.5 si a_1 et a_2 ont le même mode (par exemple, majeur). A l'inverse, il sera diminué de 0.5 si a_1 et a_2 n'ont pas le même mode.

6.1.2 Représentation des séquences d'accords

Pour assurer une certaine robustesse de la méthode de comparaison entre séquences d'accords, celle-ci se doit d'être invariante par transposition [Uit02] (la transposition est définie en section 3.3.2). De ce fait, la progression d'accords Do Majeur, Fa Majeur, Sol

Intervalle entre fondamentales (en demi-tons)	Score de substitution associé
0	+2.850
1	−2.850
2	−2.475
3	−0.825
4	−0.825
5	−0.000
6	−1.800

Table 29 – Score de substitution entre deux accords a_1 et a_2 en fonction de l'intervalle entre leurs fondamentales, d'après [HFR07].

Majeur en Do Majeur doit être considérée comme similaire à Sol Majeur, Do Majeur, Ré Majeur en Sol Majeur.

Il est ainsi possible d'utiliser à une séquence d'accords représentée de manière relative, où chaque accord est représenté par rapport à l'accord précédent (en dénombrant, par exemple, le nombre de demi-tons séparant la fondamentale de l'accord courant avec la fondamentale de l'accord précédent). Cette méthode de représentation a néanmoins été mesurée comme moins précise pour des algorithmes d'alignement local [HFR07]. En effet, si chaque accord est représenté relativement à l'accord précédent, la modification d'un accord entraine la modification de l'accords suivant, pénalisant ainsi la robustesse de la comparaison entre séquences d'accords.
Une autre manière de définir une séquence d'accords invariante par transposition est de considérer les degrés des accords par rapport à la tonalité générale. Pour ce faire, plusieurs représentations des accords relatives à la tonalité générale peuvent être envisagées. Ces différentes représentations sont nommées et décrites ci-dessous :

- représentation relative *demi-tons* : chaque accord est représenté par le nombre de demi tons séparant sa fondamentale de la tonique de la tonalité générale,
- représentation relative *cycle-quintes* : chaque accord est représenté par la distance sur le cycle des quintes entre sa fondamentale et la tonique de la tonalité générale,
- représentation relative *TPS* : chaque accord est représenté par sa distance de Lerdahl au premier degré de la tonalité générale,
- représentation relative *TPS-sym* : idem, mais en considérant une distance de Lerdahl symétrique.

Il est également possible de conserver une représentation absolue des accords, mais de prendre en compte la transpo-invariance en procédant à une adaptation dans le calcul de l'alignement local [AFHI07] pour considérer de multiples transpositions. Le principe est de transformer une des séquences d'accords dans les 12 tonalités différentes, et de considérer

la transposition minimisant les sommes des coûts pour la transformer en l'autre séquence d'accords. Dans la suite, nous considérons ainsi deux représentations absolues :

- représentation absolue *transpo* : chaque accord est représenté de manière absolue, et l'alignement local considérera les transpositions multiples,
- représentation absolue *non-transpo* : chaque accord est représenté de manière absolue, et l'alignement local ne considérera pas les transpositions multiples.

6.1.3 Évaluation et résultats

Cette section propose une évaluation du calcul de similarité, en fonction des différents paramètres (représentation des accords, score de substitution) de la méthode de comparaison entre séquences d'accords. Une comparaison avec la méthode proposée par De Haas est proposée. Cette dernière n'utilise pas d'alignement local.

Une des notions centrales dans l'évaluation de la recherche de similarité musicale est la notion de *version*. Considérons à titre d'exemple les trois morceaux présentés tableau 30.

Titre	Artiste	Album	Année
Smells Like Teen Spirit	Nirvana	Nevermind	1991
Smells Like Teen Spirit	Nirvana	From the Muddy Banks of the Wishkah	1996
Smells Like Teen Spirit	Tori Amos	Crucify	1992

TABLE 30 – Trois versions différentes d'un même titre.

Ces trois morceaux constituent trois versions différentes d'un même titre. Le premier est une version enregistrée en studio, le second, une version du même artiste (Nirvana) mais enregistrée lors d'un concert, et le troisième est une *reprise*, c'est-à-dire une version proposée par un artiste différent de l'artiste original. La notion de reprise sera détaillée en section 6.2.2. Ces trois morceaux forment une même *classe*, car ils constituent trois versions différentes d'un même titre.

6.1.3.1 Bases de données

Nous considérons deux bases de données, contenant toutes deux des morceaux de musique au format symbolique (MIDI) générés par le logiciel *Band In A Box* (voir annexe A.1.2). Pour ces deux bases de données, les annotations des séquences d'accords et de la tonalité générale sont proposées par le logiciel.

La première base de données, appelée *Jazz*, est identique à celle utilisée pour les expériences dans les travaux de De Haas en 2008 [DHVW08]. Elle est composée de 388 fichiers, représentant 242 titres de jazz différents dont 85 existent sous au moins deux versions

Représentation	1er niveau	2nd niveau	TOP 10
Relative Demi-tons	0.712	0.756	0.793
Relative Cycle-Quintes	0.722	0.766	0.796
Relative TPS	0.512	0.552	0.558
Relative TPS-sym	0.531	0.578	0.650
Absolue Non-Transpo	0.615	0.660	0.707
Absolue Transpo	**0.840**	**0.884**	**0.892**

Table 31 – Efficacité des différentes représentations des séquences d'accords pour le calcul de similarité sur la base *Jazz*.

différentes et engendrant plusieurs séquences d'accords potentiellement différentes pour le même titre. Parmi les différences entre deux séquences d'accords différentes du même titre, on citera entre autres la transposition, la différence de structure (nombre de couplets), l'omission d'accords, l'enrichissement de certains accords (comme un accord de septième enrichi en accord de neuvième) et les différences de durées.
La seconde base de données, appelée *Mix*, est composée de 578 fichiers, dont 275 classes de titres appartenant à différents styles (pop, folk, jazz, latin, etc.).

6.1.3.2 Protocole d'évaluation

Tous les morceaux issus des différentes classes de la base *Jazz* sont un à un considérés comme requête, et pour chaque requête, les autres fichiers de la base de données sont classés selon leur score de similarité. Trois descripteurs entre 0 et 1 sont calculés, comme proposé dans [DHVW08] :

- le premier niveau moyen : le nombre de versions différentes de la requête dans les meilleurs $(C-1)$ fichiers du classement, divisé par $(C-1)$, où C est le nombre de versions différentes de la requête,
- le second niveau moyen : le nombre de versions différentes de la requête dans les meilleurs $(2C-1)$ fichiers du classement, divisé par $(C-1)$,
- le top 10 : le nombre de versions différentes de la requête dans les 10 fichiers du classement, divisé par le maximum entre 10 et $(C-1)$.

Pour une requête appartenant à une classe donnée, on attend que les autres morceaux de la classe arrivent en tête du classement. Ainsi, plus ces trois descripteurs prennent des valeurs élevées, plus la recherche de similarité est considérée comme efficace.

6.1.3.3 Résultats

Représentation des séquences d'accords

Le tableau 31 présente les scores des différentes représentations possibles pour les séquences d'accords. On remarque que les représentations relatives des accords se montrent meilleures que la représentation absolue sans prise en compte des transpositions dans la méthode d'alignement (0.712 ou 0.722 pour le premier niveau contre 0.615). Néanmoins, la prise en compte des transpositions améliore de manière significative la représentation absolue (0.840). Ce score est supérieur à tous les scores des représentations relatives. Ceci peut être expliqué par certaines erreurs d'annotations de tonalité, qui sont toujours plus délicates à déterminer, et donc à annoter que les accords, surtout pour un style comme le jazz. La présence de modulations, non mentionnées par l'annotation de la seule tonalité générale, peut également expliquer ce résultat. Ceci nous conduit à adopter, pour la suite, la représentation absolue avec prise en compte des transpositions dans la méthode d'alignement local.

On notera également les mauvais scores obtenus avec la distance de Lerdahl, qui s'était pourtant montrée la plus efficace dans le système d'estimation d'accords et de tonalités locales présenté chapitre 5 (voir section 5.2.3). On peut ainsi avancer que cette distance est tout à fait adaptée à l'estimation des informations tonales, mais pas au calcul de leurs similarités.

Score de substitution

Le tableau 32 présente les résultats des différents scores de substitution présentés en 6.1.1.On note tout d'abord que, pour les scores de substitution binaires, la méthode d'alignement local se montre plus performante que la méthode décrite en [DHVW08] : 0.84 contre 0.74 pour le premier niveau. On remarque ensuite que, comme pour les résultats précédents, la distance de Lerdahl se montre la moins intéressante en terme de recherche de similarité. Les différences entre les autres scores de substitution sont peu significatifs. La meilleure définition pour ce score semble être issue de la consonance proposée dans [HFR07] et représentée dans le tableau 29. On notera enfin que la prise en compte du mode des accords permet une très modeste amélioration des résultats.

Résultats sur la base de données *Mix*

On trouvera dans le tableau 33 les résultats sur la base de données *Mix* avec une représentation des accords absolue avec prise en compte des transpositions dans la recherche d'alignement local et un score de substitution utilisant la consonance (définie tableau 29) et le mode. Ces résultats sont satisfaisants (0.685 pour le premier niveau), même s'ils se

Score de substitution	1er niveau	2nd niveau	TOP 10
Binaire	0.840	0.884	0.892
Cycle quintes	0.851	0.888	0.897
Lerdahl	0.272	0.462	0.524
Consonance	0.870	0.894	**0.904**
Consonance et Mode	**0.873**	0.894	**0.904**

Table 32 – Efficacité des différents scores de substitution pour le calcul de similarité sur la base *Jazz*.

	1er niveau	2nd niveau
Moyenne	0.685	0.732

Table 33 – Résultats du calcul de similarité sur la base *Mix*.

révèlent inférieurs à ceux issus de la base *Jazz*. Ces différences se justifient par l'absence de versions rigoureusement identiques dans la base *Mix*, contrairement à la base *Jazz*. Pour 159 morceaux (sur 275), les autres versions du titre de la requête occupent les premières places du classement de similarité.

6.1.3.4 Conclusion

La comparaison avec la méthode présentée par De Haas [DHVW08] montre que l'alignement local se révèle efficace lorsque l'on cherche à estimer la similarité entre séquences d'accords, en présentant une robustesse aux variations d'accords et en se montrant ainsi capable d'estimer comme similaires deux séquences d'accords présentant certaines différences (insertions d'accord, transpositions, enrichissements). L'adoption d'une représentation absolue des accords avec une prise en compte des transpositions dans la méthode d'alignement local permet en outre d'obtenir de meilleures performances qu'avec une représentation relative des accords à la tonalité générale du morceau. Ce résultat peut notamment s'expliquer par l'occurrence de modulations, qui n'ont pas été annotées, ou plus généralement d'erreurs dans les tonalités annotées. Enfin, il apparait que de simples scores de substitution donnent également de bons résultats. Cela tend encore à prouver que cette efficacité est due au principe général de la recherche d'alignement local, et non pas à l'affinage particulier des scores de substitution.

Les travaux présentés dans cette section ont fait l'objet de publication en conférence internationale [HRR09a].

6.2 Similarité harmonique à partir de séquences d'accord estimées

Une fois les paramètres relatifs à la comparaison de séquences d'accords définis (alignement local, score de similarité des opérations élémentaires), il est souhaitable de vérifier si la recherche de similarité peut utiliser des séquences d'accords estimées, et non plus annotés par des utilisateurs (voir figure 66). L'objectif principal de la recherche par similarité est en effet de pouvoir être applicable à toute base de données musicale, sans nécessité d'annotations préalables. Il parait donc intéressant de vérifier si la mesure de similarité reste efficace en intégrant les erreurs éventuelles engendrées par l'estimation des accords. Ainsi, le système présenté chapitre 5 sera ici utilisé pour estimer les accords des morceaux, avant de procéder à la comparaison de ces séquences d'accords.

6.2.1 Similarité harmonique de musique symbolique

Nous considérons une base de données de 1000 fichiers MIDI collectés sur Internet, et dont la qualité a été vérifiée par nos soins. Parmi ces 1000 fichiers, 850 représentent des titres uniques, et 150 représentent différentes versions de 43 titres distincts, formant ainsi 43 classes de titres (avec une moyenne de 3,49 versions différentes par classe). Comme nous l'avons mentionné précédemment, les versions d'une même classe peuvent se distinguer de différentes manières : arrangements différents, reprises, styles musicaux différents, etc. Il est à noter que dans une même classe, deux versions sont nécessairement distinctes. Le protocole d'évaluation est identique à celui présenté dans la section 6.1. La seule différence provient donc du fait que les séquences d'accords utilisées pour la comparaison ne proviennent plus d'annotations, mais sont estimées grâce au système présenté en section 5.2 (voir figure 66).

Le tableau 34 détaille les résultats de la recherche de similarité, en considérant successivement les 150 requêtes correspondant aux différentes versions des 43 classes de titres. Le premier niveau est de 0.68 et indique donc une précision intéressante : pour 19 classes de titres, toutes les autres versions de la requête arrivent aux premières places, et cela quelle que soit la requête. Plus d'un tiers (51 sur 150) des requêtes donnent les autres versions du même titre en tête de classement. Plus de 75% (116 sur 150) des requêtes voient au moins une des autres versions du même titre arriver en tête.
On note également que les erreurs proviennent de différences musicales notables entre les morceaux d'une même classe. Nous avons par exemple, et à titre de test, considéré deux versions différentes de *Viva la Vida* de Coldplay, et une version de *If I Could Fly* de Joe Satriani. Ces deux morceaux présentent une ressemblance musicale qui a récemment fait débat dans le milieu pop/rock, Satriani engageant même un recours en justice pour plagiat

1er niveau	2nd niveau	TOP 10	TOP 20
0.68	0.73	0.75	0.78

Table 34 – Résultats en utilisant le système d'estimation d'accords.

(ce recours est aujourd'hui clos, les deux parties ayant trouvé un arrangement). Le système permet de mettre en évidence une forte ressemblance entre les deux versions différentes de *Viva la Vida*, mais n'estime pas de similarité particulière entre ces deux versions et *If I Could Fly*. Ceci est explicable par la complexité des différences entre ces deux morceaux (différentes structures, différents couplets, etc).

On constate que l'intégration du système d'estimation d'accords et de tonalités présenté en 5.2 préserve la précision de l'estimation de similarité harmonique sur des séquences annotées, en présentant une efficacité comparable sur le 1er niveau et le 2nd niveau. Ces résultats ont fait l'objet d'une publication en conférence internationale [HRR09b].

6.2.2 Similarité harmonique de signaux sonores musicaux

Un des problèmes centraux de l'étude de similarité harmonique de musique audio est l'identification de reprises, ou *cover*. Cette tâche est complexe, car elle met en jeu la notion musicale de titre identique. Dans certains cas, la similitude entre l'original et la reprise est incontestable comme pour *You Really Got Me* des Kinks repris par Van Halen, ou *Hush*, de Joe South, repris par Deep Purple et ensuite par Gotthard. Dans d'autres cas, la différence est notable, comme pour *Highway To Hell* de AC/DC repris par Marilyn Manson, ou *Smells Like Teen Spirit* de Nirvana repris par Tori Amos. Ces différences peuvent induire un tempo, une orchestration, une structure différente et même un changement de langue des paroles (*Comme d'habitude* de Claude François, repris par Frank Sinatra dans *My Way*). L'identification des reprises est donc une tâche complexe [EP07].

Nous nous proposons d'utiliser un système de recherche de similarité harmonique pour l'identification de reprises en musique audio. Ce système utilise l'estimation de séquences d'accords depuis le signal audio présenté en section 5.3, et la comparaison de séquences d'accords présentée précédemment. Ce système a été proposé au concours MIREX 2010 pour la tâche *Audio Cover Song Identification*. Dans cette tâche, 1000 fichiers audio sont utilisés, parmi lesquels 30 classes comportant chacune 11 versions différentes du même titre. Chaque méthode proposée pour cette tâche est ensuite évaluée de la manière suivante : les 11 morceaux de chacune des 30 classes sont successivement considérés comme requête, engendrant ainsi 330 requêtes. On somme ensuite, pour chaque rang des 330 classements des résultats, le nombre de versions appartenant à la même classe que la requête.

	Méthodes		
Rang	MHRAF1	MOD1	**RMHAR1**
1	94	106	159
2	101	82	139
3	105	57	115
4	88	53	98
5	92	37	93
6	70	34	79
7	64	31	75
8	58	29	56
9	59	26	51
10	49	16	43
Total	780	471	**908**

TABLE 35 – Résultats du MIREX 2010 pour la tâche *Audio Cover Song Identification*. Pour chaque rang r figure le nombre de morceaux identiques à la requête ayant terminé au rang r dans les 330 requêtes considérées. La méthode proposée est la méthode RMHAR1. La méthode proposée par Martin et al. (MHRAF1) utilise un alignement local entre séquences de vecteurs chroma. La méthode MOD1 utilise une méthode d'alignement utilisant des informations rythmiques et harmoniques (détails de la méthode non disponibles à ce jour).

La méthode proposée, RMHAR1, utilise une estimation de similarité harmonique à l'aide d'une estimation préalable d'accords. Elle utilise ensuite une méthode d'alignement local en prenant en compte les transpositions. Le tableau 35 présente les résultats de la méthode proposée, comparés aux résultats des deux autres méthodes concurrentes pour cette tâche (les résultats détaillés de la tâche *Audio Cover Song Identification* du MIREX 2010 sont disponibles sur le site du MIREX [1]). On note que la méthode proposée se révèle la plus efficace en terme de recherche de similarité. Non seulement, elle totalise le plus grand nombre de morceaux appartenant à la même classe que la requête dans le top10, mais elle présente les meilleurs résultats pour chacun des huit premiers rangs de ce classement,. La méthode proposée (RMHAR1) se distingue en outre de la seconde meilleure méthode MHRAF1 (proposée par Martin et al. de l'université de Bordeaux) par le fait qu'elle utilise une estimation d'accords, là où la méthode MHRAF1 utilise le même algorithme d'alignement local, mais cette fois appliqué directement aux vecteurs chroma. L'estimation des accords, qui sont des informations tonales de plus haut niveau que les vecteurs chroma, semble donc une fois de plus bénéficier à l'estimation de similarité musicale.

1. `http://www.music-ir.org/mirex/wiki/2010:Audio\_Cover\_Song\_Identification\_Results`

6.3 Utilisation du modèle en arbre de l'harmonie

Nous avons proposé en section 3.3.1, une modélisation hiérarchique de l'harmonie. Cet arbre de l'harmonie présente un intérêt analytique : nous avons ainsi constaté que l'estimation des accords était plus précise en prenant en considérant des groupes de notes, et que l'analyse de la tonalité était plus précise en prenant en compte les accords (voir sections 4.2.1 et 4.3.1). Nous nous proposons ici d'utiliser ce modèle de l'harmonie à des fins d'estimation de similarité harmonique. Au lieu de comparer des séquences (de notes ou d'accords), nous proposons une méthode pour comparer deux morceaux de musique en comparant leurs arbres d'harmonie.

La comparaison d'arbres représente une tâche complexe, étudiée dans plusieurs domaines de recherche, comme la biologie moléculaire [CLZ00] ou la reconnaissance de motifs [Lu79]. Pour estimer une similarité entre arbres, des mesures de distance d'édition, initialement introduites pour les comparaisons de séquences de symboles [WF74], furent adaptées pour la comparaison d'arbres [Sel77, Tai79]. De la même manière que pour l'estimation de similarité entre séquences (voir section 6.1), l'estimation de similarité entre deux arbres t_1 et t_2 fait intervenir la somme des scores des opérations élémentaires transformant t_1 en t_2. Comme pour les séquences, l'ordre de ces opérations élémentaires n'a pas d'influence, de même que le sens de la transformation (de t_1 vers t_2 ou l'inverse). Le détail de la méthode de comparaison entre arbres est présenté dans [AFH+11].

Afin de mesurer l'influence de la prise en compte de l'arbre d'harmonie dans l'estimation de similarité, nous proposons de considérer certains cas de plagiats musicaux avérés, et de comparer la similarité estimée à l'aide de séquences musicales (dans ce cas, de notes ou d'accords) avec la similarité estimée à l'aide de l'arbre d'harmonie.

Nous commençons en décrivant un exemple de plagiat musical dans la section 6.3.1, avant de mesurer l'apport de l'arbre d'harmonie sur 4 exemples célèbres de plagiat dans la section 6.3.2.

6.3.1 Exemple de plagiat

Au début des années 1970, un procès s'ouvrit pour plagiat : Ronald Mack accusa George Harrison d'avoir plagié *He's So Fine* (composé par Mack en 1963 et joué par The Chiffons), avec sa chanson *My Sweet Lord* [Cop]. Bien que Harrison soutint ne pas avoir copié consciemment la mélodie de *He's So Fine*, le tribunal conclut, en 1976, qu'il avait bien (consciemment ou non) copié cette mélodie.

Pour justifier cette décision, le tribunal analysa la structure des deux chansons. La figure 68 montre deux extraits de chacune de ces deux chansons. *He's So Fine* est composé
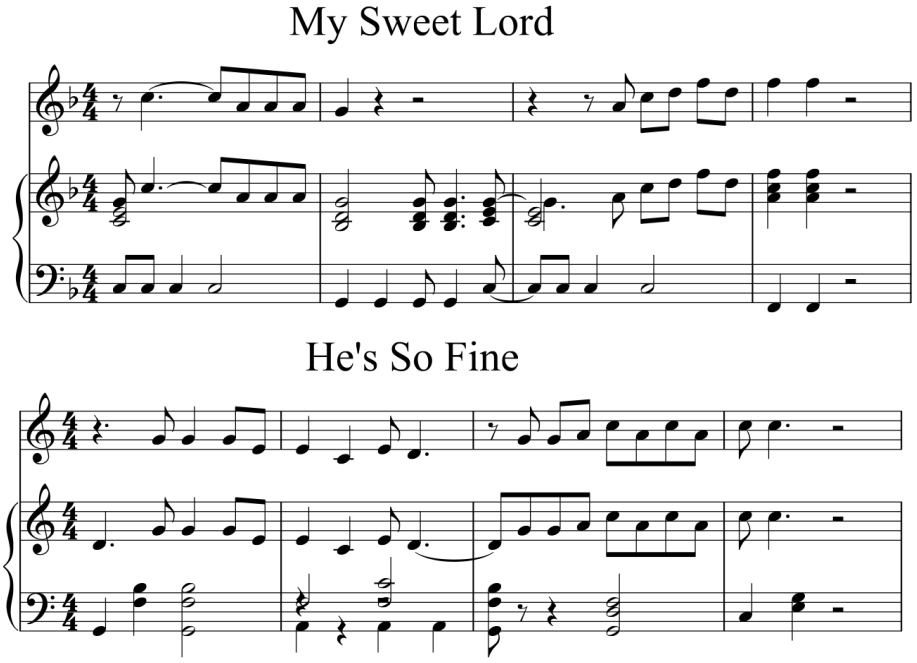


FIGURE 68 – Transcriptions manuelles d'extraits (correspondant aux motifs A et B) des deux morceaux *My Sweet Lord* (G. Harrison) et *He's So Fine* (R. Mack).

de quatre répétitions d'un court motif mélodique (motif A, sur la figure 69) suivi de quatre répétitions d'un autre motif (motif B, sur la figure 69). *My Sweet Lord* présente une structure très similaire avec quatre répétitions du même motif A, suivi de 3 répétitions du motif B. La quatrième répétition de ce motif B contient une appogiature illustrée sur le motif C. D'après ces considérations, le tribunal conclut à un cas de plagiat.
De nos jours, il est souhaitable de proposer une méthode automatique pour détecter si une pièce de musique constitue ou non un plagiat d'une autre œuvre. Nous nous proposons ainsi d'expérimenter l'efficacité de l'estimation de similarité tenant compte de l'arbre d'harmonie sur 4 cas avérés de plagiat.

6.3.2 Expériences sur 4 cas de plagiat

L'étude présentée ici concerne 4 cas de plagiat célèbres aux États-Unis lors des cinquante dernières années [Col05]. Deux morceaux distincts sont associés à chaque cas, et ont été déclarés comme "très similaires" par un tribunal. Chacun de ces morceaux, encodé au format MIDI, est successivement considéré comme une requête, et ajouté à une base de données de *bruit*, constituée de morceaux divers sans lien avec les cas de plagiat. La base de données de bruit utilisée est tirée du catalogue de chansons folkloriques Essen [2], et contient plus de 5000 pièces musicales, encodées au format MIDI. Nous attendons de

2. `http://essen.themefinder.org/`

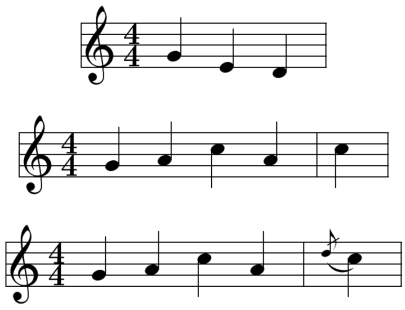


Figure 69 – Courts motifs musicaux composant la structure des deux morceaux *My Sweet Lord* (G. Harrison) et *He's So Fine* (R. Mack) : motif A (en haut), motif B (au centre) et motif C (en bas).

l'estimation de similarité non seulement qu'elle désigne comme maximum de similarité la pièce donnée en requête, mais également le cas de plagiat associé avec cette requête, en dépit des différences potentielles entre les deux morceaux constituant ce cas de plagiat.

Les morceaux considérés étant monophoniques, les groupes de notes et les notes sont confondus. De plus, les morceaux étudiés ne comportant pas de modulations, la tonalité générale et les tonalités locales sont confondues, et réduites à une simple note (la fondamentale de la tonalité). La similarité entre séquences de notes est calculée en utilisant une méthode d'alignement local proposée dans [HFR07] et dont le score de substitution entre les notes $s(n_1, n_2)$ dépend de l'intervalle entre n_1 et n_2 et est illustré tableau 29. La similarité entre séquences d'accords est également calculée à l'aide d'une méthode d'alignement local dont les scores de substitution suivent la distance sur les cycles des quintes entre les fondamentales des accords comparés (voir section 6.1). La fondamentale des accords est estimée par le logiciel *Melisma*. Le calcul de similarité entre arbres suit la méthode décrite dans [AFHI07].

Le tableau 36 montre les résultats des différentes méthodes d'estimation de similarité harmonique pour chaque requête, selon l'information tonale considérée :

- la séquence de notes,
- la séquence d'accords estimés,
- l'arbre d'harmonie.

Pour chacun de ces trois cas, nous présentons le score de similarité entre les deux pièces formant le plagiat, et le score maximal de similarité issu de la base de bruit (noté *Essen* Rang 1 dans le tableau). Pour qu'un cas de plagiat soit identifiable, il est souhaitable que le score de similarité soit significativement supérieur pour la pièce issue du même cas de plagiat que la requête, que pour le score maximal de la base de bruit.

Cette expérience permet de mettre en avant les apports de l'arbre d'harmonie au calcul de similarité entre les pièces de musique faisant l'objet d'un cas de plagiat. En effet, un

système utilisant uniquement une information sous forme de séquence (accords ou notes) ne permet pas de retrouver tous les plagiats étudiés car certains plagiats concernent une information mélodique, qui sera identifiée à l'aide des séquences de notes, mais pas à l'aide de la séquence d'accords. Pour d'autres plagiats, la ressemblance concerne davantage l'harmonie, et sera donc plus facilement mise en évidence à l'aide de l'information apportée par la séquence d'accords, et ne sera à l'inverse pas identifiable en utilisant les séquences de notes. On notera par exemple le cas de plagiat défini par *la Maritza* et *Autumn leaves*, dont la similarité estimée à partir des séquences de notes ne permet clairement pas d'établir un cas de plagiat : 11.4 contre 44.2 pour le maximum de la base de bruit (alors que ce score devrait être plus important que le maximum de la base de bruit pour pouvoir conclure à une forte similarité entre les deux pièces concernées).

En considérant le score de similarité induit par la représentation en arbre, on s'aperçoit que les cas de plagiats sont à chaque fois correctement reconnus, car ils présentent un score de similarité supérieur au score maximal sur la base de bruit.

6.3.3 Conclusion

En dépit du nombre limité de pièces utilisées pour la comparaison d'estimation de similarité, l'apport du modèle hiérarchique de l'harmonie (arbre d'harmonie) semble bénéfique. La prise en compte simultanée des séquences de notes et d'accords, lors des comparaisons entre morceaux, permet de mettre en évidence différentes natures de similarités musicales (mélodique, harmonique), et bénéficie ainsi à la précision de l'estimation de similarité. Il est maintenant souhaitable de pouvoir poursuivre de telles évaluations sur une base de données de cas de plagiats plus conséquente, afin de vérifier l'apport de ce modèle à une plus grande échelle.

Représentation	Pièce musicale Score de similarité		
R. Mack vs G. Harrison (1976)			
requête *Sweet Lord*			
	Sweet Lord	So Fine	*Essen* Rang 1
Notes	143.8	14.0	20.0
Accords	35.7	**22.7**	20.5
Arbres	179.6	**32.9**	30.0
requête *So Fine*			
	So Fine	Sweet Lord	*Essen* Rang 1
Notes	153.2	14.0	21.8
Accords	32.5	**22.7**	21.8
Arbres	187.7	**32.9**	25.3
Selle vs Gibb (1984)			
requête *Let It End*			
	Let It End	How Deep	*Essen* Rang 1
Notes	159.0	**36.3**	30.2
Accords	35.7	14.5	33.5
Arbres	194.7	**47.8**	37.3
requête *How Deep*			
	How Deep	Let It End	*Essen* Rang 1
Notes	165.0	**36.3**	28.9
Accords	39.0	14.5	27.1
Arbres	203.9	**47.8**	34.7
Heim vs Universal (1946)			
requête *Vagyok*			
	Vagyok	Perhaps	*Essen* Rang 1
Notes	200.3	**47.4**	35.6
Accords	48.7	22.9	23.3
Arbres	249.1	**57.7**	34.8
requête *Perhaps*			
	Perhaps	Vagyok	*Essen* Rang 1
Notes	170.9	**47.4**	23.2
Accords	61.7	22.9	43.6
Arbres	232.7	**57.7**	52.2
Autumn leaves vs La Maritza (1974)			
requête *Autumn Leaves*			
	Autumn	Maritza	*Essen* Rang 1
Notes	49.9	11.4	17.2
Accords	19.5	**19.5**	13.0
Arbres	69.4	**35.2**	21.9
requête *Maritza*			
	Maritza	Autumn	*Essen* Rang 1
Notes	125.3	11.4	44.2
Accords	19.5	**19.5**	13.0
Arbres	144.8	**35.2**	34.5

Table 36 – Résultats de l'estimation de similarité sur quatre cas de plagiat avérés, faisant apparaître les scores utilisant les séquences (notes et accords) et l'arbre d'harmonie.

Chapitre 7

Perspectives et conclusion

7.1 Perspectives

L'estimation des informations tonales que nous avons présentée dans les chapitres précédents laisse place à de nombreuses perspectives de travail. Nous avons ici choisi d'en distinguer quatre, et de présenter quelques résultats préliminaires.

7.1.1 Temps réel

Il est possible d'imaginer le système d'estimation d'accords présenté en 5.3 pour le temps réel et qui pourrait donc estimer en direct les accords et tonalités locales jouées par une source (musicien, orchestre, haut parleur) dont le son serait capturé par un micro. Les applications d'une telle prise en compte vont de l'accompagnement automatique en fonction du contexte tonal estimé à un instant donné, au contrôle de la performance instrumentale. On pourrait ainsi imaginer une basse et un piano virtuels accompagnant un guitariste en suivant la progression d'accords jouée par ce dernier. Sur tous les fichiers de la base de données présentée en annexe A.3, le temps de calcul de l'estimation de la séquence d'accord d'un morceau (calcul de vecteurs chroma + estimation conjointe des accords et tonalités locales) est notablement inférieur à la durée de celui-ci. Le morceau *Anna Go To Him*, par exemple, a une durée de 2 minutes et 57 secondes. Son analyse par le système présenté en 5.3, avec les réglages présentés en 5.3.3.2, requiert une analyse de 30.88 secondes, réparties en 30.45 secondes pour le calcul des différents chromagrammes, 420 millisecondes pour la constitution du graphe de candidats harmoniques, et 10 millisecondes pour la sélection finale du chemin optimal. Il semble donc possible d'envisager l'estimation des accords et des tonalités locales en temps réel du point de vue du temps de calcul.

Plusieurs adaptations seraient néanmoins nécessaires à l'intégration du calcul des accords en temps réel dans un tel système. En effet, la notion de temps réel suppose que l'estimation des accords et de tonalités au temps t de l'enregistrement doit se faire sans

connaître la musique jouée en $t + d$, avec $d > 0$. Les filtres médians centrés en t, ou l'approche multi-résolutions utilisant des informations disponibles seulement après t devront donc être modifiés. De même, pour un morceau d'une durée finale d, et analysé au temps $t < d$, le chemin optimal dans le graphe de candidats jusqu'au temps t, calculé au temps t, peut différer sensiblement du chemin optimal jusqu'au temps t calculé au temps d. A posteriori, on devra donc choisir entre conserver la séquence d'accords estimés en direct, ou choisir de relancer les calculs pour obtenir la séquence d'accords correspondant au chemin optimal pour l'intégralité de l'enregistrement.

La méthode présentée chapitre 5 semble donc pouvoir être adaptée aux contraintes de temps réel, mais non sans quelques modifications. Pour se lancer dans une telle adaptation, il serait ainsi souhaitable de débuter par le calcul en temps réel des vecteurs chroma, puis d'ajouter une à une les composantes de la méthode (énumérations des candidats, constitution partielle du graphe, sélection et affichage du chemin optimal au temps t).

7.1.2 Dictionnaire d'accords étendu

```
Dm7  CM7           CM7/6    CM7
If   I ruled the world,

       Gm7        C7     Bm7-5          E7/9
Every day would be the first day of spring,

Bm7/5  E7/9               Am            Am+7   Am7
Every heart would have a new song to sing,

          Am6          Em7        Cdim             Dm7   G7
And we'd sing of the joy every morning would bring.
```

FIGURE 70 – Une représentation des accords (au dessus des paroles) utilisés pour la chanson *If I Ruled The World*, interprétée par Tony Bennett.

Comme on peut le vérifier sur la figure 70, le dictionnaire de type d'accords généralement utilisé pour annoter des pièces de musique dépasse les seuls types majeurs/mineurs, avec l'emploi fréquent d'accords de septième et de neuvième par exemple. Certains styles, comme le jazz ou le blues, sont ainsi difficilement décrits à l'aide des seules triades majeures et mineures.

Le système présenté dans le chapitre 5 est compatible avec l'adoption d'un dictionnaire d'accords étendu. Pour l'adaptation à la musique symbolique, il est nécessaire de définir de nouvelles règles de recherche de motifs pour les nouveaux types d'accords considérés (on trouvera des exemples de règles en fonction des motifs recherchés en 5.2.2.2). Concernant l'adaptation au signal audio, il est par exemple possible de construire les vecteurs modèles d'accords correspondant aux nouveaux types d'accords considérés.

On rappelle que le vecteur modèle de la triade de Do Majeur était ainsi le suivant :

$$(1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0)$$

On peut ainsi imaginer un vecteur modèle de l'accord de Do Diminué comme étant le suivant :

$$(1, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0)$$

Néanmoins, l'intégration de ces nouveaux types d'accords risque d'introduire de nombreuses erreurs lors de l'analyse : les accords de triades majeures/mineures, généralement rencontrés en musique pop, peuvent se retrouver en minorité dans le dictionnaire d'accords étendu. Il en résulte ainsi de potentielles erreurs. Nous pouvons notamment constater cette chute de la précision dans le tableau 15, où seulement deux types d'accords (triades, et non-triades) étaient considérés et où nous avons constaté une baisse de la précision du système lors de l'estimation du types des accords. L'extension du dictionnaire d'accords pris en compte semble donc constituer une tâche délicate.

Il est cependant possible de considérer plusieurs stratégies pour intégrer de nouveaux accords sans risquer d'affecter la précision du système. Parmi ces stratégies, on distinguera :

- l'adoption d'une pondération pour les accords qui apparaissent plus rarement (le type d'accord *maj 9 13* par exemple). On peut ainsi imaginer que ces accords auront moins de chance d'être sélectionnés comme candidats, et adopter une règle qui devra être satisfaite pour qu'ils soient énumérés. Par exemple, depuis le signal audio, la valeur de corrélation entre le vecteur chroma et les vecteurs modèles d'accord devra être supérieure à un seuil. Ce seuil pourrait être fixé selon des règles musicales, ou bien par apprentissage.
- une pré-connaissance du style : selon le style musical, le dictionnaire d'accords étendu pourra être restreint aux types d'accords ayant la plus grande probabilité d'apparition dans le style courant. Ainsi, les accords de septième ont une forte tendance à apparaître dans le style blues (voir figure 72) tandis que les accords de neuvième ou de treizième ont plutôt tendance à apparaître dans des styles plus singuliers, comme le be-bop (voir figure 71).

L'adoption d'un dictionnaire d'accords étendu doit donc se faire progressivement, afin de comprendre et de restreindre les erreurs engendrées par ces considérations. Cependant, les perspectives musicales sont extrêmement importantes et intéressantes, et il sera bientôt nécessaire de considérer d'autres accords que les simples triades majeures et mineures pour étendre le champ d'application de ces recherches.
Ces questions sont actuellement au centre de l'attention des chercheurs du domaine MIR.

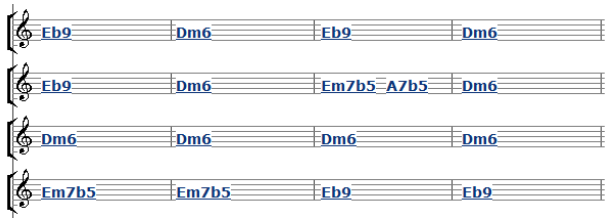


FIGURE 71 – Un extrait de la grille d'accords de *A Night In Tunisia*, de Dizzy Gillespie, appartenant au style be-bop.

```
Dm                                              D7
The thrill is gone, the thrill is gone away.
Gm                                               Dm   G Dm
The thrill is gone, baby, the thrill is gone away.
BbM7                          A7
You know you've done me wrong, baby,
                          Dm
And you'll be sorry someday.
```

FIGURE 72 – Une représentation des accords (au dessus des paroles) utilisés pour la chanson *The Thrill Is Gone*, de B.B. King, appartenant au style blues.

L'accord passé en 2010 [1] entre les participants à la tâche *Audio Chord Estimation* du MIREX présents à la conférence ISMIR, à Utrecht, propose ainsi d'adopter un dictionnaire d'accords étendu, et de progressivement abandonner la restriction aux seuls accords de triades majeures/mineures. Les signataires de cet accord ont accepté d'instaurer progressivement une évaluation tenant compte d'un dictionnaire d'accords étendu pour la tâche *Audio Chord Estimation* du MIREX.

7.1.3 Intégration des informations de structure

Certains accords répétés au cours de différents couplets, par exemple, peuvent présenter des descripteurs différents (dus à une différence de ligne mélodique au chant entre ces couplets, par exemple). Ces descriptions différentes, dans les vecteurs chroma, peuvent conduire à une estimation correcte pour un des couplets, mais erronée pour un autre. L'intégration des informations de structure pourra identifier les parties répétées, et permettre de corriger l'accord erroné en considérant qu'il est une répétition de l'accord correctement estimé.
Comme l'ont suggéré les travaux de Mauch [MND09], la prise en compte des informations de structure permet une amélioration de la précision de l'estimation des accords. Mauch utilise ainsi une détection de structure, et propose de moyenner entre elles les parties de ce chromagramme correspondant aux différentes instances d'un même motif répété (couplet ou refrain, par exemple).

1. `http://www.music-ir.org/mirex/wiki/The\_Utrecht\_Agreement\_on\_Chord\_Evaluation`

Nous proposons d'utiliser l'information structurelle (issue de [MHRF11], et appliquée au chromagramme de résolution 743 ms) de la manière suivante : si les i vecteurs chroma Cx_i sont répétitions les uns des autres, et représentent par conséquent le même accord, ils doivent partager les mêmes candidats accords. Si un accord A est candidat pour un des vecteurs chroma Cx_i, il est ainsi énuméré pour tous les vecteurs chroma Cx_i. On trouvera une illustration de cette procédure figure 73.

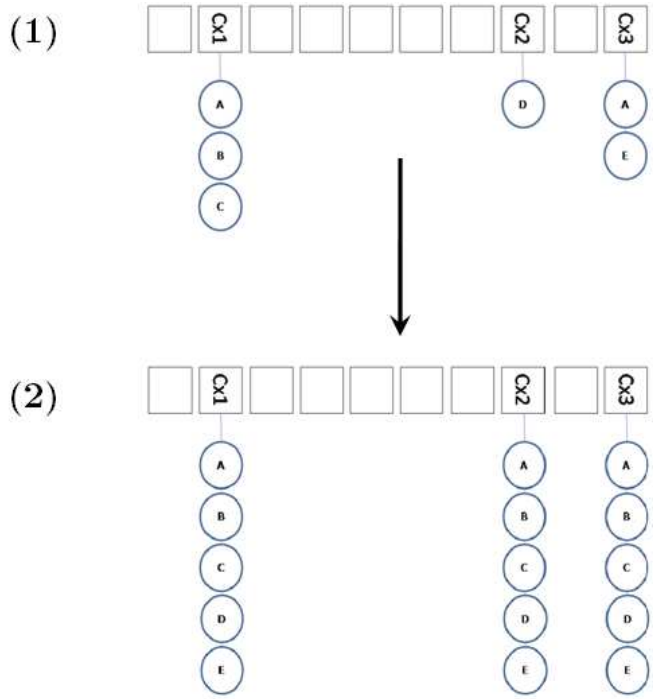


FIGURE 73 – Exemple d'utilisation des répétitions entre vecteurs chroma pour le système d'estimation d'accords. Les trois vecteurs chroma répétés Cx1, Cx2 et Cx3 peuvent partager les même candidats accords, puisqu'ils sont censés représenter le même accord.

Il est également possible d'appliquer cette mise en commun des candidats en considérant les vecteurs chroma issus de différentes tailles de trames. On propose ainsi une expérience permettant d'évaluer l'apport de plusieurs configurations :

- *total* : tous les candidats, quelle que soit leur origine, sont mis en commun (comme illustré figure 73),
- *courts* : tous les candidats issus des trames de 185 ms sont mis en commun,
- *longs* : tous les candidats issus des trames de 743 ms sont mis en commun,
- *décision* : tous les candidats issus des trames de 185 ms sont mis en commun seulement si l'énumération optimum sans prise en compte des répétitions a engendré plus d'un accord candidat.

Information structurelle	sans	avec			
Configuration	optimum	total	courts	longs	décision
Potentiel énuméré (%)	79.1	**87.7**	86.3	85.2	81.0
Nb d'accords candidats différents	1.36	2.53	2.31	2.53	1.40
Précision du système (%)	75.0	68.5	69.9	69.3	**75.8**

TABLE 37 – Potentiel énuméré, nombre d'accords candidats distincts énumérés et précision du système pour les configurations tenant compte des répétitions des vecteurs chroma.

On trouvera les résultats de l'estimation des accords sur la base Beatles dans le tableau 37. On présente également sur ce même tableau les scores relatifs à la configuration optimale obtenue en 5.3.3.2. On observe que la configuration *total* présente le meilleur potentiel énuméré, en restant inférieur à 3 candidats moyens différents par trame étudiée. Ce potentiel est meilleur qu'avec 3 candidats issus d'une même trame (voir tableau 23), confirmant l'apport d'informations pertinentes en considérant les répétitions entre les vecteurs chroma. On note également que l'ajout systématique des candidats répétés, qu'il s'agisse des candidats issus des trames de 185 ms ou de 743 ms, engendre un trop grand nombre moyen de candidats, et ne permet pas au système d'améliorer sa précision. En revanche, lorsque l'ajout des candidats répétés est conditionnel, et ne concerne que les trames pour lesquelles l'approche multi-résolutions propose plus d'un candidat, on note une amélioration, même modeste, des résultats. Ceci peut s'expliquer par le fait que lorsque l'approche multi-résolutions propose plusieurs candidats, la trame étudiée présente un certain "flou", et bénéficie de l'apport d'information extérieure, comme les candidats issus des répétitions.

La figure 74 présente la différence des scores de l'estimation d'accords obtenus entre les configurations *optimum sans structure* et *décision* sur chacun des morceaux de la base de données. On note que si la différence moyenne constatée sur le tableau 37 est faible, certains morceaux bénéficient grandement de l'apport de l'information structurelle : 6 morceaux voient même leur score augmenter de plus de 5%. Le morceau *Kansas City* connaît même une augmentation supérieure à 10% (11.8%).

Ces résultats montrent qu'une prise en compte de la structure musicale, à travers les répétitions entre vecteurs chroma, semble être prometteuse quant à l'efficacité de l'estimation d'accords pour le système proposé.

7.1.4 Utilisation de regroupements successifs de vecteurs chroma

La théorie génératrice de la musique tonale [LJ85] ou de l'analyse schenkerienne [Sch35], toutes deux présentées section 3.2, se proposent de regrouper des notes successives en *réductions* (ou *élaborations*), avant de regrouper ces réductions entre elles. Il est possible de regrouper une réduction avec une note adjacente, résultant en une nouvelle réduction. Il est

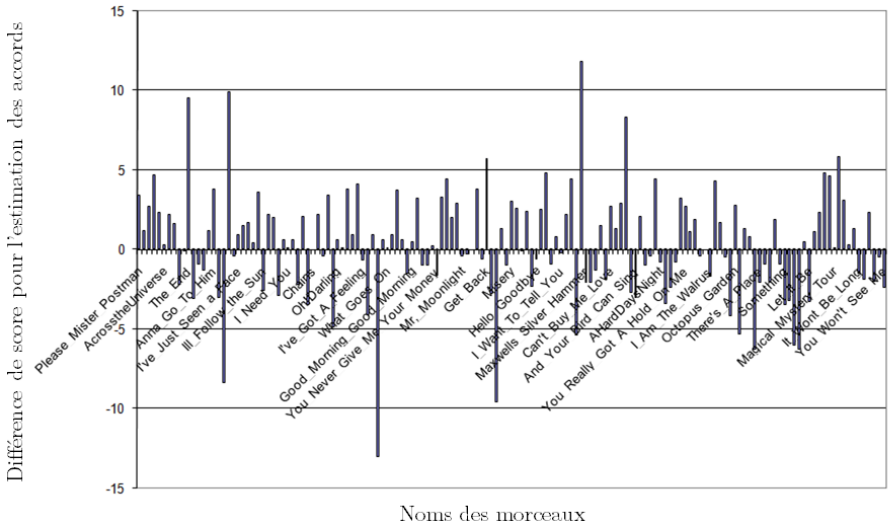


Figure 74 – Différence entre les scores de l'estimation d'accords obtenus entre les configuration *optimum sans structure* et *décision* sur chacun des morceaux de la base de données.

possible d'imaginer regrouper certains vecteurs chroma consécutifs dans le chromagramme entre eux, afin d'obtenir des objets que l'on pourrait qualifier de *méta-chromas*, et qui pourraient être regroupés à des vecteurs chroma (ou à d'autres méta-chromas) adjacents.

On propose ainsi une expérience simple, consistant à regrouper certains vecteurs chroma consécutifs selon un critère de similarité. Les méta-chromas ainsi formés ont pour valeur la moyenne des composantes respectives de vecteurs chroma les constituant. On labellise ensuite les méta-chromas par un accord (par maximum de corrélation avec des vecteurs modèle d'accords). Une fois certains accords définis, il sera possible de rattacher d'autres vecteurs chroma adjacents, selon un autre critère.

Le processus est illustré figure 75. Le critère choisi pour le regroupement entre vecteurs chroma est le suivant : on regroupe les vecteurs chroma c_i et c_{i+1} si

$$\max_j \sum_{k=0}^{35} (c_i[k].c_{i+1}[(k+j) \mod 35])$$

est atteint pour $j = \{-1, 0, 1\}$ (on conserve cette tolérance pour des raisons d'accordage, pouvant expliquer le décalage d'un indice dans deux vecteurs chroma successifs). Ce critère nous oblige à considérer des vecteurs chroma non filtrés, pour que le filtrage n'atténue pas les frontières entre regroupements de vecteurs chroma.
Concernant le critère de regroupement entre accords et vecteurs chroma adjacents, il a été choisi de regrouper un vecteur chroma adjacent à un accord si le profil modèle d'accords maximisant la corrélation avec ce vecteur chroma correspond à l'accord adjacent.

Lorsqu'aucune règle de regroupement ne peut plus s'effectuer, les vecteurs chroma isolés deviennent des méta-chromas, et sont donc labellisés par maximum de corrélation avec des vecteurs modèles d'accords.

Le protocole d'évaluation et la base de données utilisés sont semblables à ceux présentés section 5.3.3. En utilisant les règles de regroupement, on atteint une précision de 62.6% sur la base de données considérée. On remarque que cette méthode est meilleure que la méthode directe consistant à labelliser chaque vecteur chroma à l'aide de l'accord maximisant la corrélation entre le vecteur chroma et le vecteur modèle associé, qui, avec des vecteurs chroma non filtrés, obtient une précision de 58.3%.

Les regroupements successifs de vecteurs chroma adjacents semblent donc être une source d'information dont l'estimation d'accords peut bénéficier. Il semble ainsi envisageable d'adopter un pré-traitement du chromagramme pour définir ces regroupements de vecteurs chroma, avant que l'estimation des accords ne prenne place.



FIGURE 75 – Illustration des regroupements successifs de vecteurs chroma consécutifs en accords. (1) : l'entrée musicale est segmentée en vecteurs chroma.(2) : les vecteurs chroma sont regroupés en *méta-chromas*.(3) : Les *méta-chromas* sont labellisés en accords. (4) : Les frontières des accords peuvent être ajustées à l'aide de regroupement avec les vecteurs chroma adjacents.

7.2 Conclusion

Nous avons présenté, dans ces travaux, une étude approfondie des informations tonales en musique occidentale, en abordant successivement leurs définitions, leur intégration dans un modèle de l'harmonie, et leur analyse, depuis une représentation symbolique de la musique ou depuis le signal audio.

Le chapitre 2 a tout d'abord présenté les cinq informations tonales que nous avons proposées de distinguer : les notes de musiques, les groupes de notes, les accords, les tonalités locales et la tonalité générale d'une pièce de musique. Pour chacune de ces informations, une description associée à une formalisation est proposée. Nous avons choisi de lever une ambiguïté souvent rencontrée à propos de la définition de la notion d'accord, en distinguant *groupes de notes*, qui rassemblent les notes réelles jouées simultanément, et *accords*, qui réfèrent à une stabilité harmonique plus importante et décrivent un contexte tonal local dans une pièce de musique.
Deux descripteurs de l'harmonie, correspondant aux deux types de représentation informatique de la musique (représentation symbolique et signal audio) ont ensuite été présentés : les profils de notes et les vecteurs chroma. Ces deux descripteurs proposent une mesure de l'importance des classes de notes apparaissant dans un extrait musical. La durée de cet extrait, durant lequel l'observation est réalisée, est centrale : ces descripteurs peuvent en effet aussi bien servir à décrire une période de temps caractéristique d'un accord qu'à décrire une tonalité (locale ou générale).

Le chapitre 3 a ensuite présenté différentes modélisations de l'harmonie, en détaillant notamment les modèles de distances entre notes dans une tonalité donnée, et les modèles de distances entre accords. Il a ensuite été choisi de proposer un nouveau modèle de l'harmonie. Ce modèle, représenté sous la forme d'un arbre, est hiérarchique : les notes peuvent ainsi se regrouper en groupes de notes, les groupes de notes en accords, les accords en tonalités locales, et les tonalités locales en une tonalité principale, racine de l'arbre.
Un tel modèle de l'harmonie tient compte de l'influence réciproque des différentes propriétés de l'harmonie. Comme il aura été vu dans les chapitres suivants, ces influences se traduisent par des propriétés remarquables. On citera notamment la prise en compte des dépendances hiérarchiques de ce modèle pour l'analyse des informations tonales, et l'influence de ce modèle sur l'estimation de similarité musicale. Ce modèle hiérarchique est en outre original du fait que chaque niveau hiérarchique correspond à une information tonale précisément définie.

Nous avons ensuite, dans le chapitre 4, fait l'état de l'art des différentes méthodes d'analyse des informations tonales, que ce soit depuis une représentation symbolique de la musique, ou depuis le signal audio.
Nous avons également présenté différentes propositions d'amélioration de l'analyse des accords et des tonalités depuis une représentation symbolique de la musique. Concernant l'estimation des accords, nous avons ainsi montré l'influence de la prise en compte des groupes de notes : donner un poids plus important aux notes de musique composant les groupes de notes riches semble non seulement avoir un sens musical, mais permet également une amélioration de la précision de l'estimation des accords. S'agissant de l'estimation des tonalités, nous avons tout d'abord présenté une méthode d'estimation depuis les notes de musique, qui combine une approche vectorielle (associée à la distribution des notes) et matricielle (associée à la distribution des intervalles entre notes). Nous avons ensuite mis en évidence le fait que la considération des séquences d'accords permet une estimation de tonalité plus précise qu'en considérant les notes de musique.

Le chapitre 5 a présenté une nouvelle méthode d'analyse conjointe d'accords et de tonalités, applicable soit depuis une représentation symbolique de la musique, soit depuis le signal audio. Dans ce système, l'entrée musicale est découpée en segment. Pour chaque segment, des candidats accords et tonalités locales sont énumérés, puis combinés en couples candidats *(accord,tonalité)*. La programmation dynamique, associée à un calcul de distance musicale entre les candidats, permet ensuite d'extraire la séquence optimale de candidats, formant ainsi les séquences d'accords et de tonalités locales estimées.
Les adaptations de ce système aux deux différentes représentations ont été séparément décrites, puis évaluées. Les résultats depuis une représentation symbolique montrent que la précision du système proposé est comparable à celle d'un logiciel référence de l'état de l'art, et fournit, en plus de la fondamentale des accords estimés, un mode, et une séquence de tonalités locales. Les résultats depuis le signal audio prouvent que la méthode proposée est plus précise qu'une des méthodes de l'état de l'art les plus performantes pour l'estimation des accords. Ces résultats montrent également les influences bénéfiques de :

- l'approche multi-résolutions, mettant simultanément en jeu des descripteurs de l'harmonie calculés sur des périodes de temps différentes pour sélectionner différents candidats pour une même information tonale,
- la simultanéité de l'estimation des accords et des tonalités locales, en montrant que les estimations disjointes de ces deux informations perdaient toutes deux en précision par rapport à l'estimation conjointe.

Ces deux points représentent deux contributions majeures des travaux présentés ici.

Enfin, le chapitre 6 étudie l'influence des contributions présentées dans les chapitres précédents pour l'estimation de similarité musicale. Nous avons ainsi montré que les scores de similarités obtenus en utilisant des séquences d'accords estimées par le système proposé sont performants, et permettent une caractérisation des différentes versions d'un même morceau dans une base de données. La prise en compte du modèle en arbre semble elle aussi bénéficier à l'estimation de similarité, puisqu'elle peut réussir à caractériser une ressemblance musicale sur différents niveaux d'information tonale.

Dans l'avenir, le système d'estimation d'accords et de tonalités locales pourra être enrichi par l'apport d'autres considérations musicales, comme l'intégration de l'analyse de la structure, du rythme et du style. Il sera ainsi envisageable de pouvoir non seulement obtenir un arbre d'harmonie complet depuis une représentation musicale, mais de combiner ces informations tonales avec d'autres informations musicales, en présentant une interaction générale des différentes composantes de la musique dans un modèle musical complet. Les applications de tels travaux ne se limitent pas au seul domaine MIR (*Music Information Retrieval*), mais touchent les domaines de la pédagogie instrumentale, du jeu vidéo et même de l'analyse culturelle et sociale de la musique.

Publications

Chapitre de livre

Chapter : Detection of Near-Duplicate Musical Documents from a Multi-Level Comparison of Tonal Information
Pascal Ferraro, Matthias Robine, Julien Allali, Pierre Hanna, Thomas Rocher
in Information Extraction from the Internet, Nan Tang editor, iConcept Press Ltd.
ISBN : 978-0980733037 pages 129-142, 2011.

Conférences Internationales

Concurrent Estimation of Chords and Keys from Audio
Thomas Rocher, Matthias Robine, Pierre Hanna, Laurent Oudre
in Proceedings of the International Society for Music Information Retrieval Conference – ISMIR'10
pages 141-146, Utrecht, Netherlands, August 2010.

A Survey of Chord Distances With Comparison for Chord Analysis
Thomas Rocher, Matthias Robine, Pierre Hanna, Myriam Desainte-Catherine
in Proceedings of the International Computer Music Conference – ICMC'10
pages 187-190, New York City, NY, USA, June 2010.

Structured Representation of Harmony for Music Retrieval
Matthias Robine, Pierre Hanna, Thomas Rocher, Pascal Ferraro
in Proceedings of the International Computer Music Conference – ICMC'09
Montreal, Quebec, Canada, August 2009.

Dynamic Chord Analysis for Symbolic Music
Thomas Rocher, Matthias Robine, Pierre Hanna, Robert Strandh
in Proceedings of the International Computer Music Conference – ICMC'09
Montreal, Quebec, Canada, August 2009.

An Alignment Based System for Chord Sequence Retrieval
Pierre Hanna, Matthias Robine, Thomas Rocher
in Proceedings of the IEEE/ACM International Joint Conference on Digital Libraries – JCDL'09
pages 101-104, Austin, Texas, USA, June 2009.

A Robust Retrieval System of Polyphonic Music Based on Chord Progression Similarity
Pierre Hanna, Thomas Rocher, Matthias Robine
in Proceedings of the ACM's Special Interest Group on Information Retrieval, i.e. ACM SIGIR International Conference – SIGIR'09
pages 768-769, Boston, Massachusetts, USA, July 2009.

Improvements of Key-Finding Methods
Matthias Robine, Thomas Rocher, Pierre Hanna
in Proceedings of the International Computer Music Conference – ICMC'08
Belfast, Northern Ireland, August 2008.

Conférences Nationales

Analyse simultanée des accords et des tonalités depuis le signal audio
Thomas Rocher in Proceedings of the MajecSTIC (MAnifestation des JEunes Chercheurs en Sciences et Technologies de l'Information et de la Communication) – MajecSTIC
Bordeaux, France, October 2010.

Amélioration des méthodes d'estimation d'accords et de tonalité depuis une représentation musicale symbolique
Thomas Rocher, Matthias Robine, Pierre Hanna
in Proceedings of the Journées d'Informatique Musicale – JIM'09
pages 49-58, Grenoble, France, April 2009.

Bibliographie

[AFH+11] J. Allali, P. Ferraro, P. Hanna, M. Robine, and T. Rocher. Detection of Near-Duplicate Musical Documents from a Multi-Level Comparison of Tonal Information. In *Information Extraction from the Internet*, pages 129–142. iConcept Press Ltd, 2011.

[AFHI07] J. Allali, P. Ferraro, P. Hanna, and C.S. Iliopoulos. Local transpositions in alignment of polyphonic musical sequences. In Nivio Ziviani and Ricardo Baeza-Yates, editors, *14th String Processing and Information Retrieval Symposium*, volume 4726 of *Lecture Notes in Computer Science*, pages 26–38. Springer, oct. 2007.

[BA06] Alisdair MacRae Birch and Others Anon. *Band-In-A-Box Independent Users Group File Archive 2.* Available : http ://groups.yahoo.com/group/Band-in-a-Box-Files2/, 2006. [Online ; accessed 30-September-2011].

[Bel57] R. Bellman. *Dynamic Programming.* Princeton University Press, 1957.

[Bel07] J.P. Bello. Audio-based Cover Song Retrieval using Approximate Chord Sequences : Testing Shifts, Gaps, Swaps and Beats. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, pages 239–244, Vienna, Austria, September 2007.

[Bha93] J.J. Bharucha. MUSACT : A Connectionist Model of Musical Harmony. *Machine Models of Music*, pages 497–510, 1993.

[BMS00] J. P. Bello, G. Monti, and M. Sandler. Techniques for automatic music transcription. In *in International Symposium on Music Information Retrieval*, pages 23–25, Plymouth, Massachusetts, USA, 2000.

[BP05] J.P Bello and J. Pickens. Robust mid-level representation for harmonic content in music signal. In *Proceedings of International Symposium on Music Information Retrieval*, pages 304–311, London, UK, September 2005.

[BWF11] J. A. Burgoyne, J. Wild, and I. Fujinaga. An Expert Ground Truth Set for Audio Chord Recognition and Music Analysis . In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, page to appear, 2011.

[Che00] E. Chew. *Towards a Mathematical Model of Tonality.* PhD thesis, MIT Cambridge, MA, 2000.

[CLZ00] G. Collins, S. Le, and K. Zhang. A New Algorithm for Computing Similarity Between RNA Structures. In *Proceedings of the 5th Joint Conference on Information Sciences*, volume 2, pages 761–765, Atlantic City, NJ, March 2000.

[Col05] Columbia Center for New Media Teaching and Learning. *Music Plagiarism Project*, 2005. [Online ; accessed 7-October-2011].

[Cop] Copyright Website LLC. *Copyright Website.* [Online ; accessed 7-October-2011].

[Cos62] E. Costère. *Mort ou transfigurations de l'harmonie.* Presses universitaires de France, 1962.

[CV07] W. Chai and B. Vercoe. Detection of Key Change in Classical Piano Music. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, London, UK, September 2007.

[dCB03] A. de Cheveigné and A. Baskind. F0 estimation of one or several voices. In *Eurospeech*, pages 833–836, 2003.

[DEBJ10] J.S. Downie, A.F. Ehmann, M. Bay, and M. C. Jones. The music information retrieval evaluation exchange : Some observations and insights. In *Advances in Music Information Retrieval*, pages 93–115. Springer Berlin / Heidelberg, 2010.

[DHVW08] B. De Haas, R. Veltkamp, and F. Wiering. Tonal Pitch Step Distance : A Similarity Measure for Chord Progressions. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 51–56, 2008.

[DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1) :1–38, 1977.

[Dow03] J.S Downie. Music information retrieval. *Annual Review of Information Science and Technology*, 37 :295–340, 2003.

[Dow08] J.S. Downie. The music information retrieval evaluation exchange (2005-2007) : A window into music information retrieval research. *Acoustical Science and Technology*, 29(4) :247–55, 2008.

[Duc03] W. Duckworth. *A Creative Approach to Music Fundamentals.* Schirmer, 2003.

[EDB07] V. Emiya, B. David, and R. Badeau. A Parametric Method for Pitch Estimation of Piano Tones. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 249–252, Honolulu, United States, 2007.

[EP07] D.P.W. Ellis and G.E Poliner. Identifying 'cover songs' with chroma features and dynamic programming beat tracking. In *Proceedings of the International Conference on Acoustics, Speech, Signal Procesing.*, pages 1783–1794, Honolulu, Hawaii, April 2007.

[ET04] T. Eerola and P. Toiviainen. *Suomen Kansan eSävelmät. Finnish Folk Song Database.* Available : http ://www.jyu.fi/musica/sks/, 2004. [Online ; accessed 7-October-2011].

[Eul39] L. Euler. *Tentamen novae theoriae musicae ex certissismis harmoniae principiis dilucide expositae (in Latin).* Saint Petersburg Academy, 1739.

[Fin04] M. Fingerhut. Music information retrieval, or how to search for (and maybe find) music and do away with incipits. In *IAML-IASA Congress*, Oslo, Norway, August 2004.

[For73] A. Forte. *The Structure of Atonal Music.* Yale University Press, 1973.

[Fuj99] T. Fujishima. Real-time chord recognition of musical sound : a system using common lisp music. In *Proceedings of the International Computer Music Conference*, pages 464–467, Beijing, China, July 1999.

[GH04] E. Gómez and P. Herrera. Estimating The Tonality Of Polyphonic Audio Files : Cognitive Versus Machine Learning Modelling Strategies. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain, October 2004.

[Góm06] E. Gómez. *Tonal Description of Music Audio Signals.* PhD thesis, University Pompeu Fabra, Barcelona, Spain, July 2006.

[Got06] M. Goto. A chorus section detection method for musical audio signals and its application to a music listening station. *IEEE Transactions on Audio, Speech and Language Processing*, 14 :1783–1794, 2006.

[Han60] H. Hanson. *Harmonic Materials of Modern Music.* Irvington, 1960.

[HFR07] P. Hanna, P. Ferraro, and M. Robine. On optimizing the editing algorithms for evaluating similarity between monophonic musical sequences. *Journal of New Music Research*, 36(4) :267–279, December 2007.

[Hol77] S.R. Holtzman. A program for key determination. *Interface*, 6 :29–56, 1977.

[HRR09a] P. Hanna, M. Robine, and T. Rocher. An alignment based system for chord sequence retrieval. In *Proceedings of the IEEE/ACM International Joint*

Conference on Digital Libraries (JCDL), pages 101–104, Austin, Texas, USA, June 2009.

[HRR09b] P. Hanna, T. Rocher, and M. Robine. A robust retrieval system of polyphonic music based on chord progression similarity. In *Proceedings of the ACM's Special Interest Group on Information Retrieval (SIGIR)*, pages 768–769, Boston, Massachusets, USA, July 2009.

[HS05] C. A. Harte and M. B. Sandler. Automatic Chord Identification Using a Quantised Chromagram. In *Proceedings of the Audio Engineering Society*, pages 291–301, Spain, 2005.

[HSG06] C. Harte, M. Sandler, and Martin G. Detecting harmonic change in musical audio. In *Proceedings of the 1st ACM workshop on Audio and music computing multimedia*, pages 21–26, 2006.

[HTH07] K. Hirata, S. Tojo, and M. Hamanaka. Techniques for Implementing the Generative Theory of Tonal Music. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria, September 2007.

[Hub08] W. L. Hubbard. *Musical Dictionary. The American History and Encyclopedia of Music.* Elibron Classics, 1908.

[Izm05] Ö. Izmirli. Tonal similarity from audio using a template based attractor model. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, pages 540–545, London, UK, September 2005.

[Izm06] Ö. Izmirli. Audio key finding using low-dimensional spaces. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*, pages 127–132, Victoria, Canada, October 2006.

[Izm07] Özgür Izmirli. Localized Key Finding from Audio Using Nonnegative Matrix Factorization for Segmentation. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, pages 195–200, Vienna, Austria, September 2007.

[JSM93] J. Jones, D. Scarborough, and B. Miller. Gtsim a computer simulation of music perception. *Computers and the Humanities*, 27(1) :19–23, 1993.

[Kas88] M. Kassler. Apl applied in music theory. In *APL '88 : Proceedings of the international conference on APL*, pages 209–214, New York, NY, USA, 1988.

[KD06] A. Klapuri and M. Davy. *Signal Processing Methods for Music Transcription.* Springer, New York, 2006.

[KK82] C.L. Krumhansl and E.J. Kessler. Tracing the Dynamic Changes in Perceived Tonal Organisation in a Spatial Representation of Musical Keys. *Psychological Review*, 89(2) :334–368, 1982.

[Kla01] A. Klapuri. Multipitch estimation and sound separation by the spectral smoothness principle. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 3381–3384, 2001.

[Kla03] A. Klapuri. Multiple fundamental frequency estimation based on harmonicity and spectral smoothness. *IEEE Transactions on Speech and Audio Processing*, pages 804–816, 2003.

[Kla08] A. Klapuri. Multipitch analysis of polyphonic music and speech signals using an auditory model. *IEEE Transactions on Audio, Speech and Language Processing*, pages 255–266, 2008.

[KO09] M. Khadkevich and M. Omologo. Phase-change based tuning for automatic chord recognition. In *12th Int. Conference on Digital Audio Effects (DAFx-09), Como, Italy*, 2009.

[Kru90] C.L. Krumhansl. *Cognitive Foundations of Musical Pitch*. Oxford University Press, New York, 1990.

[KS79] C. Krumhansl and R. Shepard. Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology : Human Perception and Performance*, 5(4) :579–594, 1979.

[KT99] M. Karjalainen and T. Tolonen. Separation of speech signals using iterative multi-pitch analysis and prediction. In *Proceedings of the Sixth European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 2187–2190, 1999.

[KU08] P.B. Kirlin and P.E. Utgoff. A framework for automated schenkerian analysis. In *AProceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008)*, pages 363–368, Philadelphia, Pennsylvania, USA, 2008.

[Lee85] C.S. Lee. The rhythmic interpretation of simple musical sequences : Towards a perceptual model. *Musical Structure and Cognition*, pages 53–69, 1985.

[Lee06] K. Lee. Automatic chord recognition from audio using enhanced pitch class profile. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, Victoria, Canada, 2006.

[Ler01] F. Lerdahl. *Tonal Pitch Space*. Oxford University Press, 2001.

[LH78] H-C. Longuet-Higgins. The perception of music. *Interdisciplinary Science Review*, 3 :148–156, 1978.

[LH06] Y. Li and D. Huron. Melodic Modeling : A Comparison of Scale Degree and Interval. In *Proceedings of the International Computer Music Conference (ICMC)*, New Orleans, USA, November 2006.

[LHS71] H-C. Longuet-Higgins and M.J. Steedman. On interpreting Bach. *Machine Intelligence*, 6 :221–241, 1971.

[Lin92] C. Linster. On analyzing and representing musical rhythm. *Understanding music with AI : perspectives on music cognition*, pages 414–427, 1992.

[LJ85] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, Cambridge, Massachussetts, 1985.

[LS07] K. Lee and M. Slaney. A Unified System for Chord Transcription and Key Extraction Using Hidden Markov Models. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, pages 245–250, Vienna, Austria, September 2007.

[LS08] K. Lee and M. Slaney. Acoustic chord transcription and key extraction from audio using key-dependent hmms trained on synthesized audio. *IEEE Transactions on Audio, Speech & Language Processing*, 16(2), 2008.

[Lu79] S. Lu. A Tree-to-Tree Distance and its Application to Cluster Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1 :219–224, 1979.

[Mar72] J.G. Martin. Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Phychology Review*, 6(79) :487–509, 1972.

[Mar05] A. Marsden. Generative structural representation of tonal music. *Journal of New Music Research*, 4(34) :409–428, 2005.

[Mar07] A. Marsden. Automatic derivation of musical structure : A tool for research on schenkerian analysis. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR)*, pages 363–368, Vienna, Austria, September 2007.

[Mat02] B. Mathieu. Outils informatiques d'analyse musicale. Master's thesis, ENST-Bretagne, 2002.

[Max92] H. John Maxwell. *An expert system for harmonizing analysis of tonal music*, pages 334–353. MIT Press, Cambridge, MA, USA, 1992.

[MD08] M. Mauch and S. Dixon. A discrete mixture model for chord labelling. In *Proceedings of the 9th International Conference on Music Information Retrieval, Philadelphia, USA*, September 2008.

[MD10] M. Mauch and S. Dixon. Simultaneous estimation of chords and musical context from audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6) :1280–1289, 2010.

[MHRF11] B. Martin, P. Hanna, M. Robine, and P. Ferraro. Indexing musical pieces using their major repetition. In *Proceeding of the 11th annual international*

ACM/IEEE joint conference on Digital libraries, JCDL '11, pages 153–156, 2011.

[MKC05] M. Müller, F. Kurth, and M. Clausen. Chroma-based statistical audio features for audio matching. In *Proceedings ofof the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 275–278, New Paltz, USA, October 2005.

[MND09] M. Mauch, K. Noland, and S. Dixon. Using musical structure to enhance automatic chord transcription. In *Proceedings of the 10th International Conference on Music Information Retrieval, Kobe, Japan*, pages 231–236, October 2009.

[MS90] M. Mongeau and D. Sankoff. Comparison of Musical Sequences. *Computers and the Humanities*, 24(3) :161–175, 1990.

[MS00] G. Monti and M. Sandler. Monophonic transcription with autocorrelation. In *in Digital Audio Effects*, pages 257–260, Verona, Italy, 2000.

[MW07] S. T. Madsen and G. Widmer. Key-Finding with Interval Profiles. In *Proceedings of the International Computer Music Conference (ICMC)*, Copenhagen, Denmark, August 2007.

[Nat87] J.J. Nattiez. *Musicologie générale et sémiologie.* Collection Musique/Passé/Présent. C. Bourgois, 1987.

[Nol67] A.M. Noll. Cesptrum Pitch Determination. *Journal of Acoustic Society of America*, 41 :293–309, 1967.

[NW70] S.B. Needleman and C.D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3) :443 – 453, 1970.

[OBCqTG05] L. Ortiz-Berenguer, F. Casajús-quirós, and S. Torres-Guijarro. Multiple piano note identification using a spectral matching method with derived patterns. *J. Audio Eng. Soc*, 53(1/2) :32–43, 2005.

[OGF09] L. Oudre, Y. Grenier, and C. Févotte. Template-based chord recognition : influence of the chord types. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 153–158, Kobe, Japan, 2009.

[OGF11] L. Oudre, Y. Grenier, and C. Févotte. Chord recognition by fitting rescaled chroma vectors to chord templates. *IEEE Transactions on Audio, Speech and Language Processing*, 19(7) :2222–2233, 2011.

[Pau04] S. Pauws. Musical key extraction from audio. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain, October 2004.

[PB02] B. Pardo and W.P. Birmingham. Algorithms for Chordal Analysis. *Computer Music Journal*, 26 :27–49, 2002.

[PEB05] J-F. Paiement, D. Eck, and S. Bengio. A probabilistic model for chord progressions. In *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, pages 312–319, London, UK, 2005.

[Pee06] G. Peeters. Chroma-based estimation of musical key from audio-signal analysis. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*, pages 115–120, Victoria, Canada, October 2006.

[PM10] J. Pauwels and J. Martens. Integrating musicological knowledge into a probabilistic framework for chord and key extraction. In *AES Convention, 128th, Proceedings*, page 9, London, UK, 2010. Audio Engineering Society (AES).

[PP08] H. Papadopoulos and G. Peeters. Simultaneous estimation of chord progression and downbeats from an audio file. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 121–124, 2008.

[Rab90] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Readings in speech recognition*, pages 267–296, 1990.

[Ran03] D. M. Randel. *The Harvard Dictionary of Music*. The Harvard University Press, 2003.

[Rea05] *The New Real Book I : Jazz and Standards*. Hal Leonard, 2005.

[RHF07] M. Robine, P. Hanna, and P. Ferraro. Music Similarity : Improvements of Edit-based Algorithms by Considering Music Theory. In *Proceedings of the ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR)*, pages 135–141, Augsburg, Germany, September 2007.

[RIPdL06] D. Rizo, J.M. Iñesta, and P.J. Ponce de León. Tree model of symbolic music for tonality guessing. In *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, AIA 2006*, pages 299–304, Innsbruck, Austria, 2006. IASTED, Acta Press.

[Ris78] J-C. Risset. *Hauteur et timbre des sons*. Rapport IRCAM Centre Georges Pompidou, 1978.

[RK08] M.P. Ryynanen and A.P. Klapuri. Automatic transcription of melody, bass line, and chords in polyphonic music. *Computer Music Journal*, 32 :72–86, 2008.

[RK10] R.R. Randall and B. Khan. Lerdahl's tonal pitch space model and associated metric spaces. *Journal of Mathematics and Music*, 4(3) :121 – 131, 2010.

[RLM07] C. Rhodes, D. Lewis, and D. Müllensiefen. Bayesian Model Selection for Harmonic Labelling. *Mathematics and Computation in Music*, 2007.

[Roc10] T. Rocher. Analyse simultanée des accords et des tonalités depuis le signal audio. In *Proceedings of the MajecSTIC (MAnifestation des JEunes Chercheurs en Sciences et Technologies de l'Information et de la Communication) (MajectSTIC)*, Bordeaux, France, October 2010.

[RRH08] M. Robine, T. Rocher, and P. Hanna. Improvements of Key-Finding Methods. In *Proceedings of the International Computer Music Conference (ICMC)*, Belfast, Northern Ireland, August 2008.

[RRHDC10] T. Rocher, M. Robine, P. Hanna, and M. Desainte-Catherine. A survey of chord distances with comparison for chord analysis. In *Proceedings of the International Computer Music Conference (ICMC)*, pages 187–190, New York, New York, USA, June 2010.

[RRHO10] T. Rocher, M. Robine, P. Hanna, and L. Oudre. Concurrent estimation of chords and keys from audio. In *Proceedings of the International Society on Music Information Retrieval (ISMIR)*, pages 141–146, Utrecht, Netherlands, August 2010.

[RRHS09] T. Rocher, M. Robine, P. Hanna, and R. Strandh. Dynamic chord analysis for symbolic music. In *Proceedings of the International Computer Music Conference (ICMC)*, Montreal, Quebec, Canada, August 2009.

[RSC+74] M. Ross, H. Shaffer, A. Cohen, R. Freudberg, and H. Manley. Average magnitude difference function pitch extractor. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 22 :353–362, 1974.

[RUS+09] J. T. Reed, Yushi Ueda, S. Siniscalchi, Y. Uchiyama, S. Sagayama, and C.H. Lee. Minimum classification error training to improve isolated chord recognition. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 609–614, Kobe, Japan, 2009.

[Sch35] H. Schenker. *Der Frei Satz.* Vienna : Universal Edition, 1935. published in English as *Free Composition*, translated and edited by E. Oster, Longman, 1979.

[Sch38] P.A. Scholes. *The Oxford Companion to Music.* The Oxford University Press, 1938.

[Sch68] M. R. Schroeder. Period histogram and product spectrum : New methods for fundamental-frequency measurement. *Journal of The Acoustical Society of America*, 43 :829–834, 1968.

[SE03] A. Sheh and D.P.W. Ellis. Chord segmentation and recognition using EM-trained hidden Markov models. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*, pages 185–191, Baltimore, MD, 2003.

[Sel77] S.M. Selkow. The Tree-to-Tree Editing Problem. *Information processing letters*, pages 184–186, 1977.

[SGHS08] J. Serrà, E. Gómez, P. Herrera, and X. Serra. Chroma Binary Similarity and Local Alignment Applied to Cover Song Identification. *IEEE Transactions on Audio, Speech and Language Processing*, 16 :1138–1151, 2008.

[SL92] R. Sundberg and B. Lindblom. *Generative theories in language and music descriptions*. MIT Press, Cambridge, MA, USA, 1992.

[Smo79] S.W. Smoliar. A computer aid for schenkerian analysis. In *ACM 79 : Proceedings of the 1979 annual conference*, pages 110–115, New York, NY, USA, 1979.

[Ste96] R. Steblin. *A History of Key Characteristics in the 18th and Early 19th Centuries*. University of Rochester Press, New York, second edition, 1996.

[SW81] T.F. Smith and M.S. Waterman. Identification of Common Molecular Subsequences. *Journal of Molecular Biology*, 147 :195–197, 1981.

[Tai79] K. Tai. The Tree-to-Tree Correction Problem. *Journal of the Association for Computing Machinery*, pages 422–433, 1979.

[Tem99a] D. Temperley. *The Cognition of Basic Musical Structures*. The MIT Press, 1999.

[Tem99b] D. Temperley. What's Key for Key ? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered. *Music Perception*, 17(1) :65–100, 1999.

[Tem02] D. Temperley. A bayesian approach to key-finding. In *Proceedings of the Second International Conference on Music and Artificial Intelligence (ICMAI)*, pages 195–206, London, UK, 2002. Springer-Verlag.

[TK00] T. Tolonen and M. Karjalainen. A computationally efficient multipitch analysis model. *IEEE Transactions on Speech and Audio Processing*, 8 :708–716, 2000.

[Tru04] C. Truchet. *Contraintes, recherche locale et composition assistée par ordinateur*. PhD thesis, Université Paris 7 - Denis Diderot, 2004.

[TS] D. Temperley and D. Sleator. The Melisma Music Analyzer. http ://www.link.cs.cmu.edu/music-analysis/. [Online ; accessed 30-September-2011].

[TSDD93] J. Tague-Sutcliffe, J.S. Downie, and S. Dunne. Name that tune : An introduction to musical information retrieval. In *Information as a Global Commodity : Communication, Processing and Use. Proc. 21st Annual Conference of the Canadian Association for Information Science*, pages 204–216, Antigonish, Nova Scotia, July 1993. Toronto : Canadian Association for Information Science.

[Uit02] A.L. Uitdenbogerd. *Music Information Retrieval Technology.* PhD thesis, RMIT University, Melbourne, Australia, July 2002.

[UZ99] A.L. Uitdenbogerd and J. Zobel. Melodic Matching Techniques for Large Music Databases. In *Proceedings of the 7th ACM International Multimedia Conference*, pages 57–66, Orlando, USA, 1999.

[Vit67] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2) :260–269, April 1967.

[Vos00] Vos, P. G. Tonality Induction : Theoretical Problems and Dilemmas. *Music Perception, Special Issue in Tonality Induction*, 17(4) :403–416, 2000.

[VPM08] M. Varewyck, J. Pauwels, and J.P. Martens. A novel chroma representation of polyphonic music based on multiple pitch tracking techniques. In *Proceedings of the ACM International Conference on Multimedia*, pages 667–670, Vancouver, Canada, October 2008.

[Wak99] G.H. Wakefiled. Mathematical representation of joint time-chroma distribution. In *Proceedings of the SPIE Conference on Advanced Signal Processing Algorithms, Architecture and Implementation*, pages 637–645, Denver, USA, July 1999.

[WEJ09] A. Weller, D. Ellis, and T. Jebara. Structured Prediction Models for Chord Transcription of Music Audio. In *Proceedings of theInternational Conference on Machine Learning and Applications (ICMLA)*, pages 590–595, Miami, USA, December 2009.

[WF74] R.A. Wagner and M.J. Fisher. The String-to-String Correction Problem. *Journal of the association for computing machinery*, 21 :168–173, 1974.

[Wika] Wikipedia. Accord (musique). http ://fr.wikipedia.org/wiki/Accord_(musique). [Online ; accessed 30-September-2011].

[Wikb] Wikipedia. Chiffrage des accords. http ://fr.wikipedia.org/wiki/Chiffrage_des_accords. [Online ; accessed 30-September-2011].

[Wikc] Wikipedia. Circle of fifths. http ://en.wikipedia.org/wiki/Circle_of_fifths. [Online ; accessed 30-September-2011].

[Wikd] Wikipedia. Intervalle (musique). http ://fr.wikipedia.org/wiki/Intervalle_(musique). [Online ; accessed 30-September-2011].

[YL06] M-J. Yoo and I-K. Lee. Musical tension curves and its applications. In *Proceeding of the International Computer Music Conference (ICMC)*, pages 482–486, New Orleans, US, 2006.

Annexe A

Annexes : bases de données

A.1 Fichiers MIDI pour analyse des accords

A.1.1 12 morceaux pop

12 morceaux pop dont l'annotation des accords est tirée de partitions, ou recueils de chansons.

Les fichiers MIDI sont disponibles ici :
http ://dept-info.labri.fr/ñocher/PhDThesis/Pop-midi-chord.tar.gz

The Beatles - Magical Mystery Tour
The Beatles - Can't Buy Me Love
The Beatles - Ticket To Ride
The Beatles - Come Together
Otis Redding - Dock Of The Bay
Radiohead - Creep
The Eagles - Hotel California
Pink Floyd - The Wall
Pink Floyd - Comfortably Numb
Pink Floyd - Wish You Were Here
Barbara - L'Aigle Noir
The Animals - The House Of The Rising Sun

A.1.2 148 morceaux d'accompagnement

Le logiciel *Band In A Box* [BA06] est un logiciel permettant la création, l'interprétation et l'arrangement de morceau de musique au format MIDI. Il permet notamment l'accés et l'édition de la séquence d'accords des morceaux ouverts par ce logiciel. Nous avons utilisé la

base de données *Band-In-A-Box Independent Users Group File Archive 2* [BA06], mettant à disposition les pistes d'accompagnement de morceaux populaires au format Band In A Box. Les 148 morceaux d'accompagnements sélectionnés sont ainsi annotés par les utilisateurs, et leurs séquences d'accords sont connues. Ils sont divisés, toujours par les utilisateurs, en 4 styles musicaux distincts. On trouvera la liste de ces morceaux ci dessous :

Style *classique* :

ChangePartners
GypsyLife
MapleLeafRag-ScottJoplin
OdeAnDieFreude[OdeToJoy]
-LudwigVanBeethoven
OldPianoRollBlues
Progress
Rag
TheEntertainer-ScottJoplin

Style *jazz* :

4Brothers
500MilesHigh
ABancaDoDistinto
AcercateMas[ComeCloserToMe]
ACertainSmile
AChildIsBorn-ThadJones
AFelicidade
AFineRomance-Kern&Fields
AFoggyDay-GeorgeGershwin
AfricanFlower
AfternoonInParis
AfterYouveGone
AftonPaOljaren
AgainstTheWind-BobSeger
AHistoriaDeLilyBraun
Ah-LeuCha-CharlieParker
AHouseIsNotAHome-
HalDavid&BurtBacharach
Ah !SeEuPudesse
Airegin-SonnyRollins
AllMyTomorrows
AllOfMe
AllOfYou
AllOrNothingAtAll
AllTheThingsYouAre
AloneTogether
AlongCameBetty
AMeMePiaceOBlues-PinoDaniele
AmericanPatrol
AmorEmPas[OnceILoved]
AnaMaria
AnAmericanInParis
AngelEyes
AngelinaTheBaker
ANoiteDoMeuBem
AnOscar-CharlieParker
AnotherEwe
AnotherHairdo-CharlieParker
AnotherSaturdayNight-SamCooke
Anthropology-CharlieParker
APorta
AprilInParis
AprilJoy
AprilShowers
ARazzAMaTazz
Armageddon
AsiaMinor
ASongForYou
Assedic-LesEscrocs

ATouchOfBlues
AtSundown
AuClairDeLaLune
AuldLangSyne
AuPrivave-CharlieParker
AuraLee
AutumnInNewYork
AutumnLeaves
Avalon
AveMariaNoMorro
AVolta
roundMidnight-TheloniousMonk

Style *latin* :

Adios-EnricMadriguera
AikoAiko
Alegria,Alegria
AltijdIsKortjakjeZiek
ATasteOfHoney-RicMarlow&BobbyScott
BlackMagicWoman-Santana
CarelessWhisper-GeorgeMichael
Cariba-WesMontgomery
Day-O[BananaBoatSong]
Estrellita-ManuelPonce
FusionOf5Elements-MichaelHedges
GeraldinosEArquibaldos
HoedjeVanPapier
HotHotHot-BusterPoindexter
InDenHaagDaarWoontEenGraaf
InHollandStaatEenHuis
IveGotYouUnderMySkin
Jalousie
LadyOfSpain-TolchardEvans&ErellReaves
LangZalZeLeven
MalaFemmena
MinorAndDim
MourningIntoDancing
NotreSamba-EddyMarnav&GuyMagenta
OSoleMio
PazzetiniDiBikini
Rakkaudenjuhla[LoveFiesta]
RecadoBossaNova
RoadSong
SambaCarnavalesco
SambaDeSquirrel
SambaDeUmaNotaSo[OneNoteSamba]-
AntonioCarlosJobim
SoYouSay-JohnScofield
TakeTheATrain-DukeEllington[Latin]
Tendresse
TroublantBolero
Zombie

Style *pop* :

38YearsOld-TragicallyHip
40Days-RonnieHawkins
ABluesSerenade
Abraham,MartinAndJohn
AcrossTheUniverse-Beatles
ADayInTheLife-Beatles
AFellowNeedsAGirl
AfterTheGoldRush-NeilYoung
AheadByACentury-TragicallyHip
AHorseWithNoName-America
AintTooProudToBeg
AlDiLa
Alfie
AllAlongTheWatchtower-JimiHendrix
AllHaveToDoIsDream-EverlyBrothers
AllIAskOfYou-Phantom
AllInLoveIsFair
AllMyLoving-Beatles
AllOfMyLife

AllTheseThings
AlwaysOnMyMind
AmarilloByMorning-GeorgeStrait
AndILoveHer
AngelOfTheMorning-
JuiceNewton&MerileeRush
AngelsJam
Angie
Angie-RollingStones
AnotherEwe
AnywhereIWander
AsTearsGoBy-RollingStones
ATimeForLove
AWhiterShadeOfPale-ProcolHarum
CALIFDRM
CantHelpFallingInLoveWithYou-ElvisPresley
daytripper
dontGoBreaking
EternalFlame
HarddaysNight
HeyJudeBeatles
HotelCalifornia
Howdeepis
imagine
LoveStory

Ces morceaux sont annotés par des fichiers texte, ayant le même nom que les fichiers MIDI, mais présentant une extension en ".beat". Chaque ligne de ces fichiers d'annotations représentent un accord, et suit le format suivant :

temps_début durée fondamentale altération type

- **temps_début** est un entier codant le temps (ou la pulsation) auquel l'accord débute. Le premier temps du morceau est codé par un 0, le second par un 1, etc.
- **durée** est un entier codant la durée (en temps) pendant lequel l'accord perdure.
- **fondamentale** est un entier codant pour la fondamentale : 1 représente le Do, 2 représente le Ré, 3 représente le Mi, jusqu'à 7, représentant le Si.
- **altération** est un entier relatif représentant l'altération de la fondamentale : 0 pour al note naturelle, +1 pour l'altération # (dièse), -1 pour l'altération b (bémol).
- **type** est un entier codant pour le type.

Une description des types est présenté ci dessous. On trouvera à chaque fois l'entier codant pour le type, le nom de ce type, et un vecteur binaire de dimension 12 indiquant si les classes de notes sont jouées pour un accord de ce type dont la fondamentale est Do (C).

type 0
$(0,0,0,0,0,0,0,0,0,0,0,0)$

type 1 : bass
$(1,0,0,0,0,0,0,0,0,0,0,0)$

type 2 : C maj
$(1,0,0,0,1,0,0,1,0,0,0,0)$

type 3 : C 6
$(1,0,0,0,1,0,0,1,0,1,0,0)$

type 4 : C 6/9
$(1,0,1,0,1,0,0,1,0,1,0,0)$

type 5 : C add9

$(1,0,1,0,1,0,0,1,0,0,0,0)$

type 6 : C MAJ7
$(1,0,0,0,1,0,0,1,0,0,0,1)$

type 7 : C MAJ7add13
$(1,0,0,0,1,0,0,1,0,1,0,1)$

type 8 : C MAJ9
$(1,0,1,0,1,0,0,1,0,0,0,1)$

type 9 : C MAJ13
$(1,0,1,0,1,0,0,1,0,1,0,1)$

type 10 : C 7
$(1,0,0,0,1,0,0,1,0,0,1,0)$

type 11 : C 9
$(1,0,1,0,1,0,0,1,0,0,1,0)$

type 12 : C 13
$(1,0,1,0,1,0,0,1,0,1,1,0)$

type 13 : C min
$(1,0,0,1,0,0,0,1,0,0,0,0)$

type 14 : C min6
$(1,0,0,1,0,0,0,1,0,1,0,0)$

type 15 : C min6/9
$(1,0,1,1,0,0,0,1,0,1,0,0)$

type 16 : C min add9
$(1,0,1,1,0,0,0,1,0,0,0,0)$

type 17 : C min 7
$(1,0,0,1,0,0,0,1,0,0,1,0)$

type 18 : C min 7 add11
$(1,0,0,1,0,1,0,1,0,0,1,0)$

type 19 : C min 7 add13
$(1,0,0,1,0,0,0,1,0,1,1,0)$

type 20 : C min 9
$(1,0,1,1,0,0,0,1,0,0,1,0)$

type 21 : C min 11
$(1,0,1,1,0,1,0,1,0,0,1,0)$

type 22 : C min 13
$(1,0,1,1,0,1,0,1,0,1,1,0)$

type 23 : C min 7maj
$(1,0,0,1,0,0,0,1,0,0,0,1)$

type 24 : C min 9 7maj
$(1,0,1,1,0,0,0,1,0,0,0,1)$

type 25 : C min 7 (b5)
$(1,0,0,1,0,0,1,0,0,0,1,0)$

type 26 : C min 9 (b5)
$(1,0,1,1,0,0,1,0,0,0,1,0)$

type 27 : C min 11 (b5)
$(1,0,1,1,0,1,1,0,0,0,1,0)$

type 28 : C dim
$(1,0,0,1,0,0,1,0,0,0,0,0)$

type 29 : C dim7
$(1,0,0,1,0,0,1,0,0,1,0,0)$

type 30 : C dim7 add 7maj
$(1,0,0,1,0,0,1,0,0,1,0,1)$

type 31 : C aug
$(1,0,0,0,1,0,0,0,1,0,0,0)$

type 32 : C sus
$(1,0,0,0,0,1,0,1,0,0,0,0)$

type 33 : C sus 7
$(1,0,0,0,0,1,0,1,0,0,1,0)$

type 34 : C sus 9
$(1,0,1,0,0,1,0,1,0,0,1,0)$

type 35 : C sus 13
$(1,0,1,0,0,1,0,1,0,1,1,0)$

type 36 : C sus4-3 7
$(1,0,0,0,1,1,0,1,0,0,1,0)$

type 37 : C maj7 (b5)
$(1,0,0,0,1,0,1,0,0,0,0,1)$

type 38 : C maj7 (#5)
$(1,0,0,0,1,0,0,0,1,0,0,1)$

type 39 : C maj7 (#11)
$(1,0,0,0,1,0,1,1,0,0,0,1)$

type 40 : C maj9 (#11)
$(1,0,1,0,1,0,1,1,0,0,0,1)$

type 41 : C maj13 (#11)
$(1,0,1,0,1,0,1,1,0,1,0,1)$

type 42 : C 7 (b5)
$(1,0,0,0,1,0,1,0,0,0,1,0)$

type 43 : C 9 (b5)
$(1,0,1,0,1,0,1,0,0,0,1,0)$

type 44 : C 7 (#5)
$(1,0,0,0,1,0,0,0,1,0,1,0)$

type 45 : C 9 (#5)
$(1,0,1,0,1,0,0,0,1,0,1,0)$

type 46 : C 7 (b9)
$(1,1,0,0,1,0,0,1,0,0,1,0)$

type 47 : C 7 (#9)
$(1,0,0,1,1,0,0,1,0,0,1,0)$

type 48 : C 7 (b5b9)
$(1,1,0,0,1,0,1,0,0,0,1,0)$

type 49 : C 7 (#5#9)
$(1,0,0,1,1,0,0,0,1,0,1,0)$

type 50 : C 7 (#5b9)
$(1,1,0,0,1,0,0,0,1,0,1,0)$

type 51 : C 7 (#11)
$(1,0,0,0,1,0,1,1,0,0,1,0)$

type 52 : C 9 (#11)
$(1,0,1,0,1,0,1,1,0,0,1,0)$

type 53 : C 7 (b9#11)
$(1,1,0,0,1,0,1,1,0,0,1,0)$

type 54 : C 7 (#9#11)
$(1,0,0,1,1,0,1,1,0,0,1,0)$

type 55 : C 13 (b5)
$(1,0,1,0,1,0,1,0,0,1,1,0)$

type 56 : C 13 (b9)
$(1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0)$

type 57 : C 13 (#11)
$(1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0)$

type 58 : C 7 sus (b9)
$(1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0)$

type 59 : C 13 sus (b9)
$(1, 1, 0, 0, 0, 1, 0, 1, 0, 1, 1, 0)$

Les 148 fichiers avec leurs annotations, regroupés par style, sont disponibles ici :
http ://dept-info.labri.fr/r̃ocher/PhDThesis/BIAB.zip

A.1.3 57 morceaux d'accompagnement pour l'évaluation des distances

Parmi les 149 morceaux d'accompagnement précédemment décrits, 57 présentent des graphes de candidats très riches, permettant de mettre en évidence les différences entre les différents modèles de distance musicale intégrés dans le système.

Ces 57 fichiers sont disponibles ici :
http ://dept-info.labri.fr/r̃ocher/PhDThesis/BIAB-dist.zip

A.2 Fichiers MIDI pour analyse de tonalité

A.2.1 Extraits de la base d'apprentissage du MIREX

96 extraits au format MIDI dont la tonalité a été annotée et mise à disposition par le concours MIREX [DEBJ10]. Ces extraits sont polyphoniques.

Ces extrait sont disponibles ici :
http ://dept-info.labri.fr/r̃ocher/PhDThesis/Mirex-midi-keyDB.zip

Chaque fichier MIDI est accompagné d'un fichier texte ayant le même nom (mais présentant une extension en ".txt") contenant l'annotation de tonalité. Ces correspondances sont également listées ci-après :

01.mid AMaj	09.mid BMaj	17.mid C#Maj
02.mid AMaj	10.mid BMaj	18.mid C#Maj
03.mid AMin	11.mid BMaj	19.mid C#Maj
04.mid AMaj	12.mid BMaj	20.mid C#Maj
05.mid BbMaj	13.mid CMaj	21.mid DMaj
06.mid BbMaj	14.mid CMaj	22.mid DMaj
07.mid BbMaj	15.mid CMaj	23.mid DMaj
08.mid BbMaj	16.mid CMaj	24.mid DMaj

25.mid EbMaj
26.mid EbMaj
27.mid EbMaj
28.mid EbMaj
29.mid EMaj
30.mid EMaj
31.mid EMaj
32.mid EMaj
33.mid FMaj
34.mid FMaj
35.mid FMaj
36.mid FMaj
37.mid F#Maj
38.mid F#Maj
39.mid F#Maj
40.mid F#Maj
41.mid GMaj
42.mid GMaj
43.mid GMaj
44.mid GMaj
45.mid AbMaj
46.mid AbMaj
47.mid AbMaj
48.mid AbMaj
49.mid AMin
50.mid AMin
51.mid AMin
52.mid AMin
53.mid BbMin
54.mid BbMin
55.mid BbMin
56.mid BbMin
57.mid BMin
58.mid BMin
59.mid BMin
60.mid BMin
61.mid CMin
62.mid CMin
63.mid CMin
64.mid CMin
65.mid C#Min
66.mid C#Min
67.mid C#Min
68.mid C#Min
69.mid DMin
70.mid DMin
71.mid DMin
72.mid DMin
73.mid EbMin
74.mid EbMin
75.mid EbMin
76.mid EbMin
77.mid EMin
78.mid EMin
79.mid EMin
80.mid EMin
81.mid FMin
82.mid FMin
83.mid FMin
84.mid FMin
85.mid F#Min
86.mid F#Min
87.mid F#Min
88.mid F#Min
89.mid GMin
90.mid GMin
91.mid GMin
92.mid GMin
93.mid AbMin
94.mid AbMin
95.mid AbMin
96.mid AbMin

A.2.2 Chansons traditionnelles finnoises

8613 chansons folkloriques finnoises [ET04] au format MIDI, monophoniques.

Ces fichiers sont disponibles ici :
http ://dept-info.labri.fr/r̃ocher/PhDThesis/finfolk-tunes.zip

Le fichier *finfolktunes_ data.txt* contient des informations comme les transcripteurs, les signatures rythmiques des chansons, et leurs tonalités.

A.2.3 Extraits de la discographie des Beatles

86 morceaux des Beatles dont l'annotation de la tonalité générale est tirée des travaux de Alan Pollack [1]. Les fichiers MIDI (polyphoniques) sont disponibles ici :

1. http ://www.icce.rug.nl/s̃oundscapes/DATABASES/AWP/awp-notes_on.shtml

http ://dept-info.labri.fr/ř̃ocher/PhDThesis/Beatles-midi-key.tar.gz

On trouvera ci-apres la correspondance entre le nom du fichier MIDI et la tonalité du morceau correspondante :

ActNaturally : G Maj
AHardDaysNight : G Maj
All_I've_Got_To_Do : A Maj
All_My_Loving : E Maj
And_Your_Bird_Can_Sing : E Maj
Anna_(Go_To_Him) : D Maj
AnotherGirl : D Maj
AnyTimeAtAll : D Maj
Ask_Me_Why : E Maj
A_Taste_Of_Honey : F# min
Baby_It's_You : G Maj
BabysInBlack : A Maj
Because : C# min
Black_Bird : G Maj
Boys : E Maj
CantBuyMeLove : C Maj
CarryThatWeight : C Maj
Chains : Bb Maj
Devil_In_Her_Heart : G Maj
DizzyMissLizzie : A Maj
Doctor_Robert : A Maj
Don't_Pass_Me_By : C Maj
Do_You_Want_To_Know _A_Secret : E Maj
Drive_My_Car : G Maj
EightDaysAWeek : D Maj
Eleanor_Rigby : E min
Everybody's_Got_Something_To_Hide _Except_Me_and_M : A Maj
EverybodysTryingToBeMyBaby : E Maj
EveryLittleThing : A Maj
Glass_Onion : F Maj
Good_Day_Sunshine : E Maj
Good_Night : G Maj
Got_To_Get_You_Into_My_Life : G Maj
Help : A Maj
Here,_There_And_Everywhere : G Maj
HerMajesty : D Maj
Hold_Me_Tight : F Maj
HoneyDont : E Maj
Honey_Pie : G Maj
Magical_Mystery_Tour : A Maj
Martha_My_Dear : Bb Maj
MaxwellsSilverHammer : A Maj
MeanMrMustard : E Maj
Misery : C Maj
Money : E Maj
Mother_Nature's_Son : D Maj
NoReply : C Maj
Nowhere_Man : E Maj
Ob-La-Di,_Ob-La-Da : Bb Maj
OctopussGarden : E Maj
OhDarling : A Maj
Piggies : Eb Maj
Please_Mister_Postman : A Maj
Please_Please_Me : E Maj
PolythenePam : E Maj
Revolution_1 : A Maj
RockAndRollMusic : A Maj
Rocky_Raccoon : G Maj
Roll_Over_Beethoven : D Maj
Run_For_Your_Life : D Maj
Savoy_Truffle : E Maj
Sexy_Sadie : G Maj
SheCameInThroughTheBathroomWindow :

D Maj
Something : C Maj
SunKing : D Maj
Taxman : D Maj
TellMeWhatYouSee : G Maj
TellMeWhy : D Maj
TheNightBefore : G Maj
There's_A_Place : E Maj
The_Word : G Maj
TicketToRide : A Maj
Till_There_Was_You : F Maj
Twist_And_Shout : A Maj
Wait : F# min
What_Goes_On : E Maj
WhatYoureDoing : G Maj
Why_Don't_We_Do_It_In_The_Road : D Maj
Wild_Honey_Pie : G Maj
WordsOfLove : A Maj
Yesterday : F Maj
YouCantDoThat : G Maj
YouLikeMeTooMuch : G Maj
You_Really_Got_A_Hold_On_Me : A Maj
YoureGoingToLoseThatGirl : E Maj
You_Won't_See_Me : A Maj

A.3 Enregistrements studio des Beatles

174 morceaux des Beatles, enregistrés en stéréo et échantillonnées à 44100 Hz. Chacun de ces morceaux est annoté, selon la séquence d'accords et de tonalités locales, par un fichier texte. Ces fichiers sont disponibles sur le site suivant :
http ://www.isophonics.net/content/reference-annotations-beatles

Album : Please Please Me

I Saw Her Standing There
Misery
Anna (Go To Him)
Chains
Boys
Ask Me Why
Please Please Me
Love Me Do
P. S. I Love You
Baby It's You
Do You Want To Know A Secret
A Taste Of Honey
There's A Place
Twist And Shout

Album : With the Beatles

It Won't Be Long
All I've Got To Do
All My Loving
Don't Bother Me
Little Child
Till There Was You
Please Mister Postman
Roll Over Beethoven
Hold Me Tight
You Really Got A Hold On Me
I Wanna Be Your Man
Devil In Her Heart
Not A Second Time

Money

Album : A Hard Day's Night

A Hard Day's Night
I Should Have Known Better
If I Fell
I'm Happy Just To Dance With You
And I Love Her
Tell Me Why
Can't Buy Me Love
Any Time At All
I'll Cry Instead
Things We Said Today
When I Get Home
You Can't Do That
I'll Be Back

Album : Beatles for Sale

No Reply
I'm a Loser
Baby's In Black
Rock and Roll Music
I'll Follow the Sun
Mr. Moonlight
Kansas City- Hey, Hey, Hey, Hey
Eight Days a Week
Words of Love
Honey Don't
Every Little Thing
I Don't Want to Spoil the Party
What You're Doing
Everybody's Trying to Be My Baby

Album : Help !

Help !
The Night Before
You've Got To Hide Your Love Away
I Need You
Another Girl
You're Going To Lose That Girl
Ticket To Ride
Act Naturally
It's Only Love
You Like Me Too Much
Tell Me What You See
I've Just Seen a Face
Yesterday
Dizzy Miss Lizzy

Album : Rubber Soul

Drive My Car
Norwegian Wood (This Bird Has Flown)
You Won't See Me
Nowhere Man
Think For Yourself
The Word
Michelle
What Goes On
Girl
I'm Looking Through You
In My Life
Wait
If I Needed Someone
Run For Your Life

Album : Revolver

Taxman
Eleanor Rigby
I'm Only Sleeping
Love You To
Here, There And Everywhere
Yellow Submarine
She Said She Said
Good Day Sunshine
And Your Bird Can Sing
For No One
Doctor Robert
I Want To Tell You
Got To Get You Into My Life
Tomorrow Never Knows

Album : Sgt. Pepper's Lonely Hearts Club Band

Sgt. Pepper's Lonely Hearts Club Band
With A Little Help From My Friends
Lucy In The Sky With Diamonds
Getting Better
Fixing A Hole
She's Leaving Home
Being For The Benefit Of Mr. Kite !
Within You Without You
When I'm Sixty-Four
Lovely Rita
Good Morning Good Morning
Sgt. Pepper's Lonely Hearts Club Band (Reprise)
A Day In The Life

Album : Magical Mystery Tour

Magical Mystery Tour
The Fool On The Hill
Flying
Blue Jay Way
Your Mother Should Know
I Am The Walrus
Hello Goodbye
Strawberry Fields Forever
Penny Lane
Baby You're A Rich Man
All You Need Is Love

Album : The Beatles - CD1

Back in the USSR
Dear Prudence
Glass Onion
Ob-La-Di, Ob-La-Da
Wild Honey Pie
The Continuing Story of Bungalow Bill
While My Guitar Gently Weeps
Happiness is a Warm Gun
Martha My Dear
I'm So Tired
Black Bird
Piggies
Rocky Raccoon
Don't Pass Me By
Why Don't We Do It In The Road
I Will
Julia

Album : The Beatles - CD2

Birthday
Yer Blues
Mother Nature's Son
Everybody's Got Something To Hide
Except Me and My Monkey
Sexy Sadie
Helter Skelter
Long Long Long
Revolution 1
Honey Pie
Savoy Truffle
Cry Baby Cry
Revolution 9
Good Night

Album : Abbey Road

Come Together
Something
Maxwell's Silver Hammer
Oh ! Darling
Octopus's Garden
I Want You
Here Comes The Sun
Because
You Never Give Me Your Money
Sun King
Mean Mr Mustard
Polythene Pam
She Came In Through
The Bathroom Window
Golden Slumbers
Carry That Weight
The End
Her Majesty

Album : Let It Be

Two of Us
Dig a Pony
Across the Universe
I Me Mine
Dig It
Let It Be
Maggie Mae
I've Got A Feeling
One After 909
The Long and Winding Road
For You Blue
Get Back