

Orchestration musicale dynamique

Philippe ESLING

Sous la direction de
Carlos AGON

Remerciements

Je tiens à remercier Carlos Agon qui par son accueil, son soutien constant et indéfectible, ses conseils précieux ainsi que sa bonne humeur m’a permis de réaliser ce stage dans les conditions les plus agréables possibles. Je tiens également à remercier Gérard Assayag pour son accueil dans l’équipe Représentation Musicales et pour ses avis éclairés tout au long de mon stage. Un grand merci à Karim “Columbo” Haddad pour sa relecture des parties musicologiques du rapport ainsi que pour sa constante présence et sa jovialité permanente. Je remercie Grégoire Carpentier pour avoir pris de son temps pour incorporer le modèle temporel dans *Orchidée* ainsi que pour son aide indispensable et sa patience au fil des problèmes rencontrés.

Je tiens à remercier Georges Bloch pour sa relecture et ses conseils précieux ainsi que pour l’entretien qu’il nous a accordé, nous permettant ainsi de nous orienter sur la voie de la modélisation temporelle dans l’orchestration. Un grand merci à Mikhaïl Malt, Alain Banquart et Yan Maresz pour les entretiens qu’ils nous ont accordés sur leurs visions respectives de l’orchestration. Je remercie spécialement Hugues Dufourt qui nous a accueillis chez lui pour nous accorder un entretien qui fut particulièrement riche en enseignements et en réflexions pour la suite de notre travail.

Je remercie Juan José Burred pour son aide et son expertise sur la séparation de sources ainsi que pour la mise à disposition de ses travaux. Merci également à Geoffroy Peeters pour avoir pris de son temps pour orienter notre recherche sur les modèles temporels instrumentaux. Un grand merci à Axel Roebel pour son expertise et ses réponses rapides à de nombreux problèmes. Merci également à Xavier Rodet pour m’avoir accordé un entretien sur les modèles statistiques instrumentaux. Je tiens également à remercier Chungsin Yeh pour nos discussions sur la transcription musicale ainsi que ses conseils avisés. Je tiens à remercier Arshia Cont pour son aide sur la compréhension des projections parcimonieuses ainsi que la mise à disposition de son travail.

Un grand merci à tous mes confrères du master ATIAM présents à l’IRCAM, Arnaud Dessein pour nos discussions passionnées et inspirantes et Baptiste Bohelay pour sa présence joyeuse et sa motivation permanente. Merci également aux autres confrères d’ATIAM qui étaient malgré la distance toujours parmi nous.

Je remercie particulièrement mes parents pour leur soutien et leurs conseils avisés tout au long de mon parcours. Enfin, je ne remercierai jamais assez Justine, celle qui par son soutien, sa patience et sa beauté m’a permis de toujours garder confiance et continuer à avancer.

Contents

1	Introduction	1
2	Vers un système d'orchestration dynamique	3
2.1	D'hier à demain	3
2.1.1	De l'empirisme passé	4
2.1.2	Au traité contemporain ?	4
2.1.3	Orchestration et instrumentation	5
2.1.4	Ecriture et orchestration	5
2.1.5	Différents points de vue sur l'orchestration	6
2.2	Le timbre comme vocabulaire	6
2.2.1	Multiplicité du timbre	6
2.2.2	La musique précède la science	8
2.2.3	La variabilité du timbre instrumental	8
2.2.4	Le timbre orchestral	9
2.2.5	Les phénomènes d'émergence	9
2.2.6	Le timbre comme vocabulaire	10
2.3	Le temps comme syntaxe	10
2.4	Orchestrer par l'informatique	11
2.4.1	A la croisée des mondes	11
2.4.2	Systèmes existants	12
2.4.3	Discussion	12
2.5	Notre approche	13
3	Modèle micro-temporel	15
3.1	Modèles existants	15
3.2	L'incertitude d'un modèle	16
3.3	Diriger l'incertitude	16
3.4	Microvariations fréquentielles	17
3.4.1	Transformée à Q constant	17
3.5	Microvariations temporelles	19
3.5.1	Approximation polynomiale	19
3.5.2	Modèles de mixtures gaussiennes	19
3.5.2.1	Distribution gaussienne	19
3.5.2.2	Modèles de mixtures gaussiennes	20
3.5.2.3	Estimation du maximum de vraisemblance	20
3.5.2.4	Estimation par Expectation Maximisation	21
3.5.2.5	Variance libre	21
3.5.2.6	Variance fixe	24
3.5.3	Modification de durée	25
3.6	Résultats	26
3.6.1	Base de données	26

3.6.1.1	Extension de la base de données	27
3.6.2	Comparaison	27
3.6.2.1	Précision	27
3.6.2.2	Complexité calculatoire	28
3.6.3	Intégration dans <i>Orchidée</i>	28
3.6.3.1	Expériences	29
3.6.3.2	Comparaison	29
4	Evolution macro-temporelle	33
4.1	Factorisation de matrices non-négatives	33
4.1.1	Philosophie générale	33
4.1.1.1	Propriétés fondamentales	34
4.1.2	Applications	34
4.1.3	Différentes fonctions de coût	34
4.1.4	Détermination des matrices	35
4.1.4.1	Algorithmes multiplicatifs	36
4.1.4.2	Algorithmes de descente de gradient	36
4.1.4.3	Algorithmes des moindres carrés alternés.	36
4.1.5	Extensions	37
4.1.5.1	Contraintes simples	37
4.1.5.2	Factorisation parcimonieuse	37
4.1.5.3	Factorisation locale	38
4.1.5.4	Factorisation discriminante	39
4.1.5.5	Factorisation pondérée	39
4.2	Notre approche	40
4.3	Résultats	41
4.3.1	Validation	41
4.3.2	Critiques	42
5	Conclusion	43
5.1	De la complexité de l'orchestration	43
5.1.1	Complexité combinatoire	43
5.1.2	Complexité timbrale	44
5.1.3	Complexité temporelle	45
5.2	Axes émergents	46
6	Annexes	48

Chapter 1

Introduction

Nos travaux ont porté durant ce stage sur la réalisation d'un système informatique d'aide à l'orchestration. Jusqu'au 19^{ème} siècle, les compositeurs ont utilisé de manière extensive la hauteur, l'intensité et la durée comme paramètres de l'expressivité musicale. Mais il fallut attendre l'époque des compositeurs romantiques tels que Hector Berlioz, pour que le timbre devienne un paramètre en soi et que l'orchestration prenne pleinement sa place dans le discours musical. Dans son traité [6], Berlioz construisit les bases de l'orchestration musicale. Dès lors, celle-ci pouvait être considérée comme l'art subtil du mélange des sons instrumentaux et deviendrait par la suite une composante centrale de la puissance expressive de la musique. De nos jours, les compositeurs contemporains utilisent le timbre comme un élément essentiel de l'esthétique musicale. Une importante somme d'information sur l'orchestration réside dans les différents traités. On peut ainsi citer celui de Rimski-Korsakov [56] ou encore l'impressionnant ouvrage en 4 volumes de Koehlin [37]. Cependant même dans les travaux les plus récents comme ceux de Piston [54] ou Casella [15], le problème de l'orchestration a toujours été adressé par les compositeurs d'une manière empirique. Comme il n'existe pas de règles fixes et génériques, comme celles qui régissent l'harmonie dans la musique occidentale, les possibilités de combinaisons instrumentales sont toujours réduites à des exemples et conseils qui ne permettent pas d'imaginer une utilisation scientifique de tels traités.

La jeunesse de ses principes et la difficulté d'établir un formalisme adéquat pourrait expliquer pourquoi la littérature scientifique concernant ce sujet est si parcimonieuse. En effet, même si beaucoup d'attention a été portée sur les aspects symboliques de la composition, la dimension de la composition orchestrale reste l'un des derniers sujets inexplorés par l'informatique musicale. En effet l'orchestration est un problème non trivial qui s'inscrit au croisement des axes de complexité combinatoire, timbrale, temporelle et esthétique. Toutefois, les progrès en traitement du signal et les récentes percées dans la connaissance du timbre et de l'analyse musicale ont permis un regain d'intérêt dans l'étude des aspects scientifiques de l'orchestration. Plusieurs systèmes ont été mis au point [36, 53, 58, 14] reposant sur l'idée d'un *son cible* qui représente l'intention du compositeur en ce qui concerne les caractéristiques de timbre. Compte tenu de cet objectif, indiqué par un son enregistré, le système doit trouver un mélange de sons instrumentaux dans une base de données qui correspond au mieux à la cible. Cependant, tous les travaux antérieurs sur ce sujet ont mis l'accent sur une orchestration *verticale*, partant de l'analyse ciblée sur la partie entretenue des instruments pour trouver la meilleure combinaison se rapprochant de la cible, ignorant ainsi complètement l'évolution temporelle des sons instrumentaux. Toutefois, Helmholtz [34], même limité par ses moyens d'investigation, avait déjà remarqué en 1863 que la reconnaissance d'un son est liée à son évolution temporelle plutôt qu'au seul spectre de sa partie stationnaire. En outre, la notion de timbre en lui-même n'existe que dans la dimension du temps puisque les propriétés spectrales prennent naissance par l'évolution temporelle des ondes. Il convient également de noter que, dans sa pratique quotidienne de l'orchestration, le compositeur ne travaille pas uniquement par superposition verticale de sons stationnaires, mais bien plutôt par une sculpture simultanément verticale et horizontale du son. Après des entretiens avec de nombreux compositeurs, il est apparu indispensable d'aborder le problème du temps dans l'orchestration musicale. Ainsi parmi tous les problèmes soulevés par ce sujet,

nous avons décidé de nous concentrer sur la temporalité ce qui permet de nous inscrire dans la continuité des développements entrepris en passant à un niveau supérieur de modélisation.

Après une analyse de l'orchestration et de sa position dans le savoir musical au chapitre 2, nous dégageons deux modulations du temps que nous appellerons *micro-temporelle* et *macro-temporelle*, le domaine du micro-temps étant relatif aux qualités du timbre et celui du macro-temps désignant l'organisation du discours musical. Nous étudierons donc l'orchestration musicale à travers le prisme de ces deux modulations. Dans le chapitre 3, nous examinerons les modèles permettant d'appréhender le timbre en tant que structure spectro-temporelle. Nous proposons ainsi un modèle d'évolution temporelle des propriétés spectrales basé sur les mixtures de gaussiennes que nous comparons et validons par rapport à ceux existants. Ce modèle permet en outre de capturer les propriétés pertinentes des timbres instrumentaux avec une grande réduction dimensionnelle tout en permettant une resynthèse efficace de ceux-ci. De plus, le point fort de ce modèle est de permettre d'inférer de manière quasi automatique la structure temporelle d'un son identique mais de durée modifiée tout en conservant les segments d'attaque et de relâchement. Nos travaux s'inscrivant dans la continuité des développements entrepris, nous présentons les résultats d'expériences issues de l'intégration de nos travaux dans le système d'orchestration existant *Orchidée* en notant la nette amélioration apportée au modèle statique.

Après la définition du modèle micro-temporel, nous discutons dans le chapitre 4 d'une avancée possible dans l'échelle du temps vers le niveau du macro-temps. Cependant, l'idée d'une continuité entre macro et micro-temps doit sûrement être prise avec la plus grande précaution en raison de l'inexistence d'un lien direct entre ces différentes échelles temporelles. Ainsi notre recherche s'engageant dans la voie du temps, il nous faut abandonner, ne serait-ce que temporairement, l'approche combinatoire et verticale de l'orchestration pour se concentrer sur une modélisation du temps. Nous choisissons donc une première approche pour tenter d'étendre l'orchestration par un son cible en permettant à celui-ci d'être un morceau complet ayant différents éléments à orchestrer. En inversant le paradigme, l'orchestration musicale peut être considérée comme un problème de séparation de sources où le but est de trouver la combinaison d'instruments qui explique au mieux la cible. Compte tenu de cette approche, nous utilisons la factorisation en matrices non-négatives pour permettre de décomposer le spectre de la cible en une somme pondérée des instruments choisis. L'objectif est d'abord de découper le spectre en un ensemble d'unités temporelles qui ont une certaine cohérence spectrale. Cette technique permet à la fois de trouver le mélange instrumental 'optimal' pour un élément mais également d'envisager l'utilisation du système dans un cadre de *réorchestration* d'une musique entière. Ainsi grâce à la technique de factorisation de matrices non-négatives nous pouvons calculer rapidement la décomposition sur une base instrumentale. D'une part, l'ajout des contraintes de parcimonie nous permet de limiter l'effectif instrumental, d'autre part les contraintes de continuité locale nous permettent également d'ajouter une logique temporelle en pénalisant les propositions ayant de trop grandes variations instrumentales (à la fois de notes et de classes d'instruments). Nous concluons finalement en dégageant un ensemble de directions et questions ouvertes par l'orchestration suivant ses différents axes de complexité.

Chapter 2

Vers un système d'orchestration dynamique

2.1 D'hier à demain

L'orchestration est l'art d'écrire pour l'orchestre, de combiner les sonorités propres à chaque instrument en tenant compte de leurs possibilités. L'origine de cet art se situe dans la volonté des compositeurs à élever l'orchestre au rang d'unité expressive, servant au mieux un texte mis en musique ou retranscrivant le plus fidèlement certaines émotions. En premier lieu, il est effectivement question d'un *art* à part entière. Certes l'orchestration requiert bon nombre de compétences techniques, sur lesquelles nous reviendrons, mais pour une grande part, elle appelle au génie, celui qui fait l'artiste. Ainsi que Berlioz le souligne [6] à propos de l'art d'orchestrer, celui-ci "*s'enseigne aussi peu que celui de trouver de beaux chants, de belles successions d'accords et des formes rythmiques originales et puissantes*". Où nous pouvons définir un chant, une succession d'accord et une forme rythmique, il n'est pas concevable d'interpréter le beau, ou l'original. De ce fait, de toutes les composantes de l'écriture musicale, l'orchestration est longtemps demeurée, dans son enseignement comme dans sa pratique, une activité empirique. La difficulté de trouver un formalisme adéquat et rigoureux ainsi que la jeunesse de son propos dans le discours musical font que l'orchestration reste l'une des dernières dimensions musicales qui n'ait pas été traitée par l'informatique musicale à la mesure de son ampleur et de sa complexité. Cet art revêt pourtant d'une importance capitale étant donné que toute pièce comportant plus d'un instrument, tel un duo à cordes, fait appel à une idée d'orchestration. Ainsi, la mélodie peut être vue comme un croquis, une esquisse de la volonté musicale qui ne pourra être complétée que par l'ajout de couleurs par le biais des timbres instrumentaux. Il suffit alors de considérer l'éventail de timbres couvert par chaque instrument pour se faire une idée de l'étendue des possibilités sonores offertes par un orchestre. Nous pourrions dire de l'orchestration qu'elle est l'ensemble des techniques d'écriture dans lesquelles le timbre agit comme paramètre principal. Walter Piston [54] écrivait dans son traité "*il s'agit de découvrir comment l'orchestre est utilisé pour traduire la pensée musicale. [...] C'est un moyen d'étudier la manière dont les instruments se mélangent pour créer un équilibre de sonorités.*"

Aussi, au delà du problème de l'orchestration "classique", nous pouvons entrevoir une généralisation de cette question en la manière dont les sons se mélangent pour produire une unité musicale agréable. Les théories éculées de l'harmonie ne suffisent plus à satisfaire le besoin de connaissance sur les mélanges sonores. Les ruptures technologiques du siècle dernier appellent à une compréhension plus globale de ce que nous appellerons l'*harmonie timbrale*. Il faut alors se tourner vers le passé pour envisager l'avenir de l'orchestration.

2.1.1 De l'empirisme passé ...

Dans son traité, Berlioz [6] introduisit les bases de ce que deviendrait l'orchestration musicale. S'en suivit une série de traités qui même par leur nombre restreint permettent d'accéder à une somme d'information considérable. L'ouvrage le plus riche et complet à tout niveau d'exigence serait celui de Charles Koechlin [37], qui par ailleurs, du haut de ses quatre tomes, tient véritablement de l'encyclopédie. Cependant même dans les travaux les plus récents comme ceux de Piston [54] ou Casella [15], la question de l'orchestration a toujours été adressée par les compositeurs d'une manière empirique. Par l'absence de règles fixes et génériques, comme celles qui régissent l'harmonie dans la musique occidentale, les possibilités de combinaisons instrumentales sont toujours réduites à des exemples et conseils qui ne permettent pas d'imaginer une utilisation scientifique de tels traités. Ainsi dans les traités les plus récents disponibles (vieux d'une soixantaine d'années), on peut lire que la flûte est un instrument aux nuances typiquement pastorales, ou qu'il sied au basson et à ses tons burlesques un rôle de vieux libidineux. Des règles rigoureuses et "universelles" ne pouvant être exposées, que peut donc faire l'auteur d'un traité d'orchestration, sinon produire une collection d'exemples appropriés tirés du répertoire, et en discuter l'effet orchestral obtenu par une verbalisation subjective ? Les différents traités se ramènent ainsi à une série de "recettes", repérant dans le répertoire classique des archétypes orchestraux.

Pour appuyer cette approche empirique de l'orchestration, nous pourrions prendre ici l'exemple du début de la pièce "*Le sacre du printemps*" de Igor Stravinsky. L'introduction fut originellement écrite pour le cor anglais mais le soir de la première, Stravinsky entendit le basson jouer la mélodie en coulisse et fut frappé par la sonorité de celui-ci et il décida donc à la dernière minute de remplacer l'affectation de sa ligne mélodique¹. Il n'avait donc pas trouvé la "meilleure" orchestration et sa rencontre fortuite permit d'ajouter la tension et l'atmosphère inquiétante qu'il recherchait originellement. Nous remarquons ici qu'un phénomène de hasard n'est pas à exclure de la réussite d'une orchestration.

Le compositeur lorsqu'il est amené à orchestrer est alors confronté au choix d'utiliser son expérience personnelle ou de se tourner vers les traités d'orchestration qui commencent à dater et manquent d'exploration et d'expérimentation sur les combinaisons possibles du timbre. Car aussi riches d'exemples et exhaustifs que soient les traités, la question de leur pertinence et de leur portée mérite toutefois d'être posée. Quelle est leur utilité pour un compositeur aujourd'hui ? Et sans même discuter l'éventuelle obsolescence du répertoire d'étude, le vocabulaire employé ne reflète-t-il pas déjà la vision esthétique de son époque ? Il ne semble pas qu'un traité d'orchestration sur des bases scientifiques puisse exister à ce jour, cependant sa nécessité reste cruciale.

2.1.2 Au traité contemporain ?

Au vu de cette tradition empirique, nous pouvons nous demander alors s'il est possible d'approcher l'orchestration d'une manière scientifique. Comme Bregman [9] l'entrevoit déjà, "*il devrait être possible de rédiger un traité d'orchestration énonçant des principes assez abstraits pour être appliqués à tous les styles de musique. Ce ne serait pas un recueil de préceptes accumulant impératifs et interdits, mais une sorte de guide indiquant la marche à suivre pour obtenir tel ou tel son, et laissant toute liberté au compositeur de décider du résultat souhaité.*"

La nécessité épistémologique d'un traité d'orchestration scientifique reste évidente pour beaucoup de compositeurs car, notamment avec la musique électroacoustique, s'est opéré un revirement dans la musique instrumentale, qui tend de plus en plus à dépasser les catégories classiques de l'écriture (hauteur, durée, nuance) et s'attache à l'inharmonique, la composition de sons bruités et de sons à forte variation de timbre dans le temps. Les marques les plus évidentes de cette évolution sont la *musique spectrale* (e.g. Grisey, Murail) et la *musique concrète instrumentale* (e.g. Lachenmann, Sciarrino) ainsi que les écoles de la complexité (e.g. Ferneyhough, Dillon) qui cherchent à "saturer l'écriture". Remarquons que ce problème est propre à la musique occidentale, enfermée dans son écriture, puisque les musiques orales et d'autres cultures abordent depuis longtemps ces catégories du son complexe et bruité. Parmi les compositeurs d'aujourd'hui, beaucoup sont résolument tournés vers les qualités expressives du son, vers le potentiel des sons complexes, bruités ou de source électronique, comme

¹Alain Banquart, communication personnelle.

matériau de base. L'orchestration dans le cadre de ces sonorités et techniques de composition devient donc une opération cruciale. Ce problème nous ramène alors de manière plus globale à la question du rapport entre son et bruit. Comment peut-on définir la frontière entre le son et le bruit, entre la musique et le bruit ? Avec le fleurissement incessant de la musique électronique, ne sommes-nous pas amené à repousser cette limite dans ses derniers retranchements ? L'avènement de la technologie nous a ainsi fait considérer des sons jusqu'alors bruités comme étant musicaux, en témoigne la démarche de Pierre Schaeffer. Celui-ci choisit d'enregistrer et de combiner des sons produits par des sources quelconques issues de notre environnement quotidien. En effet, il considère [59] qu'il faut "*remplacer la variété restreinte des timbres des instruments que possède l'orchestre par la variété infinie des timbres des bruits obtenus au moyen de mécanismes spéciaux*". Cette démarche est confrontée au notre irrépressible réflexe d'identification d'une source sonore. Schaeffer propose pour remédier à ce problème de transformer les sons, les répéter ou de modifier l'écoute en la rendant réduite ou acousmatique.

Il semble donc que notre écoute soit dirigée par nos attentes et que ce soit la logique et la dépendance causale (par les relations d'antériorité) qui définissent la frontière entre la musique et le bruit, notre conception musicale étant liée à notre type d'écoute. On peut donc craindre que l'art de mélanger les timbres reste très empirique et le jugement qualitatif des couleurs d'orchestre typiquement subjectif. La seule vérité, aujourd'hui, est à trouver du côté de l'instrumentation.

2.1.3 Orchestration et instrumentation

Il y a lieu de faire ici le distinguo entre l'orchestration et l'instrumentation, lesquelles sont souvent confondues. L'instrumentation concerne l'étude des instruments, de leur tessiture, de leurs modes de jeux, de leurs potentialités sonores, de leur jouabilité, de leurs ressources de dynamique, d'articulation, de vitesse ainsi que leur place et leur fonction dans l'orchestre, éventuellement de leur timbre, et plus rarement de leur capacité à traduire individuellement une idée musicale. C'est un préliminaire incontournable pour l'étude de l'orchestration. Pour Piston [54], "*on n'insistera jamais assez sur l'importance de la connaissance instrumentale. Une écriture appropriée aux instruments est sans aucun doute le facteur déterminant dans la réussite d'une orchestration.*"

On peut catégoriser les composantes de l'orchestre à plusieurs échelles : de celle de l'instrument (violon, violoncelle, clarinette, flûtes) à la famille instrumentale (cordes, bois, cuivres) en passant par le mode de production. On se rend compte, ne serait-ce qu'à une échelle d'étude, de l'immensité du catalogue qu'imposerait un bilan exhaustif de l'ensemble de l'instrumentarium classique. Par exemple, si on s'intéresse à la gestuelle des instruments, la même mélodie aura une difficulté inversement proportionnelle entre la guitare et le violon. Au lieu de la méthode traditionnelle qui consiste à aller chercher dans un texte la vérification d'un élément technique, il semble qu'il serait important de réaliser enfin une aide à l'instrumentation moderne dans laquelle l'ordinateur répond aux questions de l'utilisateur. Cela est rendu possible par le fait que l'instrumentation dans sa pratique peut être assimilée à une approche de la programmation par contraintes. En effet, l'instrumentation est généralement réalisée a posteriori, l'idée génératrice étant confrontée à la réalité mécanique de sa propre réalisation.

L'orchestration, quant à elle, renvoie aux multiples manières de mélanger ces timbres individuels et ainsi "*étudier de quelle façon ils peuvent concourir à l'effet musical dans leur association*" [6]. L'orchestration est donc un art d'un niveau de complexité supérieur à l'instrumentation, et qui, d'une certaine façon, la contient : "*l'orchestration succède à l'instrumentation. La conception trop répandue que l'orchestration n'est que l'attribution de timbres aux diverses lignes nous semble très inadéquate.*" [37]. On voit ici le lien fort qui existe entre écriture et orchestration qui résultent souvent d'un processus indivisible.

2.1.4 Écriture et orchestration

Il est extrêmement difficile de séparer la fonction du compositeur de l'orchestrateur, en d'autres termes un matériau musical est souvent pensé et écrit pour un instrument. Une aide à l'orchestration ne pourrait s'inscrire que dans un moment ponctuel de l'écriture, touchant au domaine des possibilités techniques

instrumentales. Excepté la situation de l'accord à orchestrer, situation ponctuelle où le paramètre de hauteur est stable, les autres contextes musicaux (figures, textures, gestes) semblent relever conjointement de l'écriture et de l'orchestration. Précisons d'ailleurs que, dans sa pratique quotidienne d'orchestrateur, le compositeur ne travaille pas seulement en superposant verticalement des sons stationnaires mais bien plutôt en ciselant, verticalement et horizontalement, le son. Il simule l'attaque par un instrument, la tenue par un autre, la résonance par un troisième, et réalise même instrumentalement des catégories telles que l'évolution dans l'espace du son.

Deux modes d'écriture se distinguent pour l'orchestration. Ainsi il existe une écriture directe pour l'orchestre où la mélodie et l'instrument sont pensés comme une unité indivisible. D'un autre côté, un type d'écriture "abstraite" consiste à écrire la partition puis orchestrer celle-ci. On voit ainsi deux modes d'orchestration :

- Orchestration projective : écrire une partition et décider de l'affectation des différentes lignes mélodiques aux instruments. Même si Koechlin décriait cette pratique, nous pouvons prendre ici l'exemple de Maurice Ravel qui réécrivant ses pièces pour piano n'avait pas "imaginé" l'orchestre.
- Orchestration inductive : avoir un timbre en tête et chercher quelle combinaison d'instruments (ainsi que leurs hauteurs et modes) le réalise.

Ainsi le processus d'orchestration est directement lié à l'écriture, voire indivisible de la composition car il n'est pas isolé du processus compositionnel. Cependant, la problématique de l'écriture musicale est un terrain délicat pour la recherche scientifique qui touche là une technique dont l'appréciation est soumise à un jugement esthétique, et dont les mécanismes paraissent difficilement formalisables.

2.1.5 Différents points de vue sur l'orchestration

Il n'existe pas un *type* d'orchestration. Chaque musique relève d'une approche personnelle de l'orchestration et il est évident qu'une orchestration différente implique également une musique différente. De la même manière le rôle et la façon d'aborder l'orchestration sont directement liés à la personne qui l'emploie. On peut donc dégager, des traités et entretiens avec des compositeurs, différents points de vue sur l'orchestration. Tout d'abord, la position de l'orchestration par rapport à l'ensemble du savoir musical varie selon les époques et les auteurs. Koechlin, par exemple, semble la considérer [37] comme une technique totalement asservie aux autres dimensions de l'écriture : *"Il faut planifier avec soin, par rapport à l'œuvre dans son ensemble, les éléments orchestraux suivants : accents, cadences, progressions dynamiques, sommets, chevauchements. Les changements de timbre sont fonction du contexte musical"*.

D'après Hugues Dufourt, on peut voir l'orchestration sous deux angles qui sont complémentaires. D'un point de vue timbral, l'orchestration est de prime abord l'art de gestion des volumes instrumentaux. D'un autre côté, en s'intéressant aux aspects temporels de l'orchestration, on peut l'entrevoir comme un processus de gestion de l'inertie musicale². Les possibilités de combiner différents instruments pour obtenir l'idée qui n'existe que dans l'imaginaire du compositeur sont donc fortement liées à celui-ci. Notons par ailleurs qu'il existe un métier d'orchestrateur dont la discipline révèle toute son utilité dans le cas de la réorchestration, soit la réécriture d'une pièce pour un effectif instrumental différent.

2.2 Le timbre comme vocabulaire

2.2.1 Multiplicité du timbre

Lorsque le timbre entre en jeu dans le discours musical, le premier problème qui surgit est celui de sa définition même. Qui tente de formuler une définition juste du timbre se heurte inmanquablement à un certain nombre de difficultés, tant l'usage systématique du mot a fini par le rendre hermétique et insaisissable. L'American Standards Association définit le timbre comme *"les attributs de la sensation en termes desquels un auditeur peut juger que deux sons de même hauteur et même volume sonore"*.

²Hugues Dufourt, communication personnelle.

sont différents". Le timbre serait donc une propriété intrinsèque à l'instrument qui interviendrait dans les tâches de différenciation mais ne serait pas quantifiable positivement pour autant. A l'origine, le timbre désignait la corde en boyau mise en double au-dessous de la caisse d'un tambour pour mieux le faire résonner. Ainsi dès le départ, le timbre désigne une entité fixée, immuable et indissociable de l'instrument et qui lui confère une "couleur" particulière.

On trouve ainsi dans l'*Harmonielehre* de Schoenberg [60] que *"la hauteur devient perceptible grâce à la couleur sonore dont une des dimensions est la hauteur. La couleur sonore est donc la catégorie principale, la hauteur, une subdivision. La hauteur n'est pas autre chose que la couleur sonore mesurée selon une direction."* Le timbre semble donc englober la hauteur mais il peut également gêner la perception mélodique. C'est l'antagonisme entre écoute causale et écoute qualitative. On peut donc entrevoir deux facettes du timbre émerger suivant ces types d'écoute. D'un côté l'aspect causal du timbre, celui de l'origine sonore qui a produit ce qui nous est donné à entendre. D'un autre côté, les changements de sonorité que l'on pourrait qualifier de "couleurs" du son. Même si ces deux types d'écoute se complètent, ils mettent en oeuvre deux processus cognitifs différents.

Le timbre et les musiciens

Berlioz en introduisant son traité [6] déclare que *"l'objet de cet ouvrage est donc d'abord, l'indication de l'étendue et de certaines parties essentielles du mécanisme des instruments, puis l'étude fort négligée jusqu'à présent, de la nature du timbre, du caractère particulier et des facultés expressives de chacun d'eux et enfin celle des meilleurs procédés connus pour les grouper convenablement"*.

Ainsi lorsqu'on parle de timbre instrumental, nous devons prendre en compte le caractère particulier de chaque instrument ainsi que ses facultés expressives. Dans l'histoire de la musique occidentale, la voix a toujours été considérée comme l'instrument majeur et les instruments, autrefois considérés comme profanes, sont devenus progressivement des "voix" dont l'identité sonore favorisait l'écoute polyphonique. Dans le même temps, ils étaient employés en tant que sonorités spécifiques fournissant des variations de couleur sonore au discours musical. Dans ce contexte, il apparaît ainsi régulièrement que les deux facettes complémentaires du timbre fassent toujours partie intégrante de la création musicale : celui de l'identité d'un instrument reconnaissable et celui du changement de sonorité soit par variation de jeu soit par association de plusieurs instruments. Finalement, considéré du point de vue de l'écriture musicale, le timbre joue un double rôle selon les deux pôles de l'univers instrumental qui d'après Pierre Boulez sont l'articulation et la fusion. L'articulation du discours musical se fonde sur la fonction identitaire du timbre, la fusion (ou l'illusion) qui opère l'impossibilité d'identification, vise à orienter l'écoute de l'auditeur vers de nouveaux éléments du discours musical.

Le timbre et les acousticiens

La première des études importantes ayant été faites sur le timbre au 19^e siècle est due au grand physicien Hermann von Helmholtz [34] en 1863. Mais alors limité par ses moyens d'investigation, il est dans l'obligation de restreindre son étude à la partie stable des sons, ce qu'il nomme le "timbre musical", et suggère que le timbre est uniquement défini par sa composition harmonique. Il note déjà cependant *"un peu de réflexion nous fait voir à présent, que quelques-unes de ces particularités dépendent de la façon dont les sons commencent et finissent."* A la suite d'Helmholtz, les acousticiens ont continué à étudier le timbre, en faisant l'impasse sur les données temporelles. Cette approche est aujourd'hui largement remise en cause. Car malgré les études successives, il semble n'y avoir aucun consensus unique sur une définition opérationnelle ou constitutive du timbre à partir de laquelle les chercheurs pourraient construire des méthodes ou modèles empiriques. On retrouve encore la dualité sémantique du timbre. Le timbre causal qui est en jeu dans la reconnaissance de la source. De l'autre côté, la sonorité ou couleur sonore qui permet de qualifier le son produit. Le timbre est donc à la fois l'identité de la source sonore et les qualités sonores qu'elle possède. Le timbre est donc le "matériau" primaire du son. En musique les sources sont habituellement connues, il s'agit des instruments de musique. Cependant, le XX^e siècle amène l'aventure des "sons nouveaux", des musiques expérimentales, suscitées par les possibilités de synthèse et d'enregistrement du son qui ont véritablement fait émerger le problème du timbre comme axe central de recherche en informatique musicale. Dans cette ère nouvelle, l'écoute

semble plus orientée sur le *son* que sur la *musique*. Il s'agit alors de donner aux bruits le statut de sons "musicaux" avec des gammes, hauteur, intensité, durée. Ces nouvelles démarches appellent à un approfondissement de notre connaissance du timbre pour arriver à suivre et accompagner les nouvelles orientations musicales.

2.2.2 La musique précède la science

A partir du *VI^e* siècle, la musique fait partie du *quadrivium* au côté de l'arithmétique, la géométrie et l'astronomie et elle est donc considérée comme une science mathématique. Même si de nos jours, la musique est plus associée à l'art qu'aux sciences, la démarche d'analyse de Helmholtz nous a permis de comprendre que la composition en harmoniques des sons diffèrent, dans l'amplitude de chacun de ces harmoniques. Les spectres d'amplitude présentant de telles dissemblances d'un instrument à un autre, il en a légitimement été déduit que le timbre était simplement déterminé par la composition en harmoniques d'un son. Il s'avère que cette conception physique du son musical suit et coïncide avec les styles baroques, classiques et le début du romantisme de l'histoire de la musique, styles qui privilégient une hiérarchie harmonique stricte, comme par exemple, celle que théorise fortement Jean-Philippe Rameau dans ses divers traités d'harmonie : de bas en haut s'étagent la basse, fondement de l'harmonie, puis les parties intermédiaires d'accompagnement, puis la mélodie à l'aigu. On voit là une forte analogie : la basse comme son fondamental, les parties intermédiaires comme premiers harmoniques naturels, et la mélodie comme une émanation harmonique supérieure. La musique semble ainsi précéder (et accompagner) la conception scientifique du son. Ainsi comme le disait Friedrich Nietzsche, "*ce n'est pas la théorie qui englobe l'art, c'est l'art qui englobe la théorie en tant que fiction*".

Ces découvertes scientifiques suivent, et coexistent, avec une préoccupation des compositeurs de plus en plus marquée pour le timbre, et une conception musicale qui consiste à former un substrat sonore et une évolution musicale à partir d'éléments caractérisés, de courts motifs, qui se combinent ou se succèdent pour créer une forme plus globale. L'orchestre est ainsi démultiplié en petites entités musicales indépendantes mais qui, réunies, fusionnent pour façonner une enveloppe d'ensemble. C'est là que revient poindre notre analogie, les éléments musicaux qui agissent comme des partiels d'un son, créant une enveloppe sonore que l'on perçoit globalement. Encore une fois, la musique précède la science. La musique occidentale est largement basée sur ce "paradigme harmonique", c'est-à-dire qu'elle construit ses concepts musicaux à partir de sons considérés comme harmoniques et stationnaires. L'Occident a ainsi élaboré des organologies réduisant les inharmonicités et transitoires d'un son (colophone des cordes et archet, embouchure des cuivres). La notation musicale occidentale est d'ailleurs, sans hasard, cohérente avec ce paradigme (en notant un point en guise de "hauteur", on réduit un son à son fondamental, plus exactement à sa tonie, et l'on considère que dans un son, les évolutions des harmoniques sont parallèles à celle du fondamental), tout comme le sont les systèmes musicaux (polyphonie, tonalité). La science a aussi largement adopté ce paradigme. Les théories scientifiques s'intéressent principalement au son harmonique et stationnaire, on peut se demander si une avancée scientifique ne nous amènerait pas à reconsidérer l'inharmonique et le bruiteux comme partie intégrante de notre perception musicale et permettant de nouveau à la science par l'art de l'orchestration d'accompagner la musique dans ses nouvelles évolutions.

Cependant, il n'est pas d'orchestration possible sans une connaissance approfondie des instruments et de leurs capacités sonores. Mais peut-on concevoir un formalisme qui puisse capturer les propriétés acoustiques des instruments ? On ne peut répondre à cette question sans en passer au préalable par la compréhension du timbre instrumental.

2.2.3 La variabilité du timbre instrumental

Même à travers un mode de jeu particulier, les caractéristiques des sons résultants d'un instrument sont d'une variété presque infinie. Cette combinatoire presque illimitée se décline suivant les différents modes d'écoutes. L'écoute s'opère sur les notes et intervalles (mélodies), les rapports de durées (rythmes) et le matériau sonore (catégorie / qualité sonore des sources). Cette variabilité des sons instrumentaux est un obstacle difficile pour l'étude du timbre (c'est le dilemme de l'acousticien). Il existe

ainsi de nombreuses dimensions de variabilité du timbre instrumental qui peut être observée sur de multiples échelles. On peut cependant souligner trois dimensions fondamentales de la variabilité, à savoir la tessiture, la dynamique et la position spatiale. La tessiture représente l'ensemble des notes disponibles pour un instrument, la dynamique désigne le champ de variation du volume sonore et enfin la position spatiale modifie les conditions de réception du signal par un auditeur. Afin de modéliser les propriétés à la fois spectrales et temporelles des instruments, nous avons à faire face à toutes ces dimensions de la variabilité et au problème de la nature d'une certaine manière stochastique de sons instrumentaux. Ainsi nous devons garder à l'esprit que, même à hauteur, dynamique et position spatiale identiques, le timbre peut varier d'un enregistrement à l'autre. C'est pourquoi la pertinence d'une approche avec une base de données de sons enregistrés, est discutable. En effet, l'échantillon instrumental n'est qu'un exemple de réalisation possible du timbre associé à celui-ci. En outre, les échantillons dans la base de données ne devraient pas être normalisés (qui est souvent le cas) et doivent être bien calibrés pour tenir compte de la capacité dynamique de l'instrument. Un modèle instrumental complet devrait donc être fondé sur plusieurs échantillons de la même note jouée à la même intensité. Toutefois, cela paraît pratiquement impossible en raison de la diversité des conditions d'enregistrement d'une base de données à l'autre et parfois même au sein de la même base de données.

2.2.4 Le timbre orchestral

Le tournant de la seconde moitié du XX^e siècle, avec l'avènement technologique nous a permis de nous familiariser avec de nouveaux timbres. Les instruments électroniques qui apparaissent alors, plus que de renouveler l'effectif orchestral, introduisent des timbres "inouïs", profilant ainsi une nouvelle dimension dans l'imaginaire sonore. Le langage musical va alors progressivement s'étoffer avec les pratiques nouvelles que constituent les modes de jeu. Ces nouveautés font maintenant partie intégrante du discours musical contemporain. Les instruments produisent des sons et le rôle du compositeur consiste à les combiner.

Depuis le XIX^e siècle, l'orchestre est peu à peu vu comme une gigantesque palette sonore avec laquelle composer. Le timbre, jusque là conçu comme une propriété intrinsèque des instruments, va progressivement s'en détacher pour devenir dans un premier temps l'élément d'un nouveau langage musical, jusqu'à être pensé comme le pivot central de l'écriture. Il s'agit alors de créer de nouveaux timbres, de "composer le timbre" en puisant dans les richesses multiples du vocabulaire instrumental. Le grand nombre d'instruments et la diversité des timbres offrent une palette quasi infinie de couleurs orchestrales. Car de même qu'existe le timbre d'une note isolée, on peut parler de timbre orchestral pour un mélange instrumental.

2.2.5 Les phénomènes d'émergence

C'est d'ailleurs dans ce contexte, lorsqu'ils prennent naissance du mélange des sons, que l'on peut percevoir plus clairement les phénomènes d'émergence. Ces phénomènes sont encore très mal compris de nos jours et sont pourtant implicitement inclus dans la question de l'orchestration. En effet, on ne peut séparer le problème de l'orchestration du problème du timbre. Chaque instrument permet de produire un éventail de timbres qui lui sont propres. Malgré cela deux instruments créent un autre timbre que la simple addition de leurs timbres respectifs. Les rapports entre instruments créent donc des timbres différents et les phénomènes d'émergence peuvent provoquer l'apparition dans le spectre de composantes n'appartenant à aucun des instruments d'origine.

Le phénomène le mieux compris de nos jours est celui de la phase. Ce phénomène simple peut même provoquer l'annulation complète du son lors d'une quadrature de phase. Ainsi la phase a été longuement étudiée et elle semble profondément impliquée [19] dans la reconnaissance des instruments. D'autres phénomènes intéressants peuvent émerger d'une accumulation de sources acoustiques comme le couplage, d'où prend naissance le phénomène d'unisson. Inversement, on peut également voir les phénomènes de masquage et de fusion instrumentale comme émergents même si ceux-ci provoquent une disparition perceptive de composantes.

2.2.6 Le timbre comme vocabulaire

Depuis Berlioz, le timbre entre dans le langage musical s'organisant en un ordre complexe qui s'érige au dessus du système instrumental. Les compositeurs superposent les unités élémentaires pour en former de plus grandes. Le timbre peut être vu comme le vocabulaire d'expressivité dans le langage musical. On pourrait alors imaginer les timbres instrumentaux comme notre alphabet et le timbre orchestral notre lexique. Pourtant, depuis des siècles, le timbre a toujours semblé échapper à la pure abstraction. Ainsi lorsqu'on parle de timbre comme vocabulaire, la verbalisation du timbre paraît attractive. Cependant, la description sémantique du timbre semble toujours inepte (sauf cas isolés comme la brillance) dans le cas d'une verbalisation libre ou lorsque de multiples adjectifs sont utilisés. On notera ainsi les résultats de la thèse d'Anne Faure [23] sur la verbalisation du timbre qui conclut par *“un non-parallélisme direct entre représentations sémantiques et représentations perceptives, ce qui va dans le sens d'une indépendance entre ces deux niveaux de représentation”*. Il semble possible de dépasser cette incohérence sémantique en considérant le timbre comme un lexique en soi, et en tentant de comprendre le rôle qu'il remplit dans le langage musical. Force est de constater un certain nombre de points communs entre musique et langage. Tous deux se déroulent dans le temps, communiquent par le canal auditif et sont constitués d'entités élémentaires dont la notation renvoie à des phénomènes sonores identifiés et reproductibles.

Ainsi on peut se demander s'il serait possible d'établir un jour des gammes et tempéraments de timbre. Si les relations entre instruments et entre timbres peuvent être explicitées, il deviendrait alors possible d'imaginer un système de raisonnement qualitatif comme quantitatif pour travailler avec le timbre comme vocabulaire d'expressivité. Ainsi les gammes et tempéraments de timbres pourraient être explicitées permettant ainsi d'établir de nouvelles stratégies compositionnelles.

Si l'on peut considérer le timbre comme le vocabulaire de l'orchestration, il devient évident que ce sont les règles syntaxiques qui nous font défaut.

2.3 Le temps comme syntaxe

Comme nous le disions, les compositeurs contemporains utilisent souvent le timbre comme élément central de l'esthétique musicale. Malheureusement, malgré cet intérêt porté au timbre, de nombreux compositeurs ne peuvent toujours pas asservir le timbre à leur volonté musicale, car le timbre prend naissance dans des phénomènes physiques compliqués et donc difficiles à décrire et à utiliser. La définition de l'American Standards Association de 1960 est ainsi assortie d'une note qui précise que *“le timbre dépend principalement du spectre du stimulus, mais il dépend aussi de la forme d'onde, la pression sonore, la fréquence de l'emplacement du spectre, et les caractéristiques temporelles du stimulus”*. Il est maintenant clair que le timbre ne dépend pas seulement de la distribution spectrale du son, mais aussi de ses nombreuses caractéristiques temporelles. Jean-Claude Risset [57] a découvert que la proportion d'harmoniques dans un spectre de cuivre augmente à mesure que l'intensité augmente. Cette caractéristique nous montre que le timbre ne peut pas être décrit par les valeurs instantanées de nombreux paramètres, mais est plutôt caractérisé par l'évolution temporelle de ceux-ci. En bref, les caractéristiques spectrales et temporelles sont inséparables.

Tout compositeur sait fort bien que le son d'une note varie en fonction du contexte, et que c'est le pattern sonore, et non la note isolée, qui détermine la perception. De plus, le timbre n'existe que par le temps. C'est la temporalité qui permet aux ondes de se déplacer et de coexister à des fréquences différentes. Pour atteindre une syntaxe du timbre, il est nécessaire d'explicitier les relations temporelles, les processus polyphoniques ainsi que l'enchaînement des phrases musicales. L'orchestration permettrait de proposer des outils d'agencement et de distribution verticale et horizontale de sons. Par exemple, une des marques les plus prégnantes de l'utilisation du temps dans l'orchestration est celle de la création de timbres par doublure attaque / entretien. Cette technique consiste à trouver la meilleure fusion dans une superposition d'un son type attaque, greffé sur une tenue ou résonance, le premier faisant office de transitoire du second. La fusion doit être entendue ici au sens strict, c'est-à-dire dans le cas présent, une doublure qui ouvre sur un nouveau timbre, sans qu'il soit permis de retrouver isolément les instruments responsables de la doublure.

Nous revenons ici à nouveau au problème de formaliser les phénomènes qui évoluent dans le temps. C'est pourquoi il semble qu'un indice de similitude timbral doive avant tout prendre en compte l'aspect temporel du timbre. Comment définir une méthode de reconnaissance d'un son par aspect temporel ? Est-il nécessaire d'utiliser des modèles gestaltistes ou de reconnaissance de formes issus des sciences cognitives ? Les méthodes temporelles et cognitivistes dans l'élaboration d'un indice de similitude sont certainement moins "propres" scientifiquement que les méthodes spectrales mais ce sont celles qui pourront présenter le plus d'innovations. Il semble donc important de poursuivre des recherches permettant de mieux comprendre, décrire, analyser, assembler et transformer les sons complexes. Le terrain est certes plus vierge que les théories déjà éprouvées sur le son harmonique, mais c'est celui qui apportera le plus d'outils aux compositeurs d'aujourd'hui.

2.4 Orchestrer par l'informatique

Les rapports entre musique et technologie ne sont pas un particularisme du XX^e siècle. L'histoire de la musique occidentale est au contraire jalonnée de confrontations avec la technique, motivées tantôt par une demande musicale, tantôt par une irruption technologique. La musique sous forme symbolique a ainsi permis la manipulation du matériau musical à un niveau d'abstraction plus élevé que la seule pratique instrumentale. Cependant l'orchestration dépasse le cadre de la symbolique et prend pied dans le monde des possibilités sonores.

2.4.1 A la croisée des mondes

Depuis les temps primitifs, les hommes ont toujours reçu l'information sonore et ont tenté de lui donner sens. Ainsi le point de départ semble toujours avoir été le signal dont l'homme a voulu extraire une symbolique représentative. La question de l'orchestration peut être traitée dans le domaine de l'écriture du timbre instrumental. Néanmoins, l'orchestration se situe à la croisée des mondes signal et symbolique. L'ordinateur offre la possibilité de manipuler un objet sonore dans un processus de composition, et simultanément d'en étudier les propriétés acoustiques au travers d'outils d'analyse. On peut ainsi lier le monde symbolique (écriture) avec le monde sonore du timbre (monde des potentialités instrumentales). Même si on peut considérer l'orchestration comme une forme de composition, par son utilisation extensive du timbre comme élément esthétique central, la portée de l'orchestration en est beaucoup plus profonde. Le son comme matériau coexiste et interagit avec des structures formelles. Il devient donc possible pour le compositeur de relier l'exploration du sonore à l'organisation de données symboliques. L'orchestration est en effet l'art de contrôler le timbre de l'orchestre à l'aide des paramètres symboliques de l'écriture musicale. Un des intérêts théoriques majeurs de l'orchestration assistée par ordinateur réside dans le recours à des représentations multi-échelles permettant la mise en relation de données hétérogènes. Se dessine ainsi un des enjeux contemporains de l'informatique musicale : l'interaction signal / symbolique. Ces deux courants de recherche longtemps restés imperméables l'un à l'autre, en partie en raison de l'hétérogénéité apparente de leurs objets d'étude. D'une part l'analyse et la synthèse numérique des sons a contribué à la compréhension du phénomène sonore et a permis la fabrication de sons jusqu'alors inouïs. D'autre part, l'approche algorithmique s'est concentrée sur les structures symboliques de l'écriture musicale. Or l'orchestration se situe précisément à l'intersection de ces deux axes de recherche ; si elle prétend créer des timbres, c'est encore par le biais de l'écriture. Elle est donc un point de rencontre incontournable entre le symbolique et le sonore. Aussi n'est-elle appréhendable par l'informatique musicale qu'à la condition d'une convergence de l'approche signal et de l'approche algorithmique. Pour rapprocher le signal de la symbolique, nous pouvons garder une arrière pensée pour l'analyse du signal pour la musicologie. Il doit y avoir des configurations spectrales et symboliques dénotant une évolution conjointe logique. Ainsi une analyse jointe entre les données spectrales et symboliques (comme la partition MIDI et l'ensemble d'instruments utilisés) permettrait de dégager les relations entre ces deux mondes.

L'orchestration est donc un processus diagonal au niveau temporel comme compositionnel. On peut le voir comme diagonal du timbral au temporel comme du signal au symbolisme. On peut considérer

l'orchestre et le rendu de celui-ci comme émergeant de ses composantes. Aristote pensait que le tout est supérieur à la simple somme des parties. Dans le cas des mélanges instrumentaux, on peut effectivement entrevoir les phénomènes de couplage acoustique qui viennent confirmer cette vision. Cependant on peut également se demander si il n'y a donc plus à entendre que le signal lui-même. Ainsi n'y aurait-il pas une description qui échappe au timbre ? Les fonctions d'analyse acoustique et de traitement des objets musicaux ne peuvent-elles pas interagir dans un environnement commun, subordonné à la pensée créatrice du compositeur ?

2.4.2 Systèmes existants

Partant de l'idée d'une cible, Psenicka [53] a développé un système appelé SPORCH (Spectral orchestration) où la recherche est basée sur un algorithme itératif de *matching pursuit*. À chaque instrument de la base de données est associé une série de hauteurs, de dynamiques et une collection de partiels perceptuellement les plus importants. La cible est analysée avec la même procédure, et l'algorithme sélectionne la meilleure correspondance instrumentale. Le spectre de l'instrument sélectionné est soustrait de la cible et retiré de la base de données. L'algorithme itère alors en essayant de minimiser la distance euclidienne entre l'entrée et le mélange précédent. Rose et Hetrick [58] ont plus tard proposé un outil permettant d'analyser les orchestrations existantes et en créer de nouvelles, toujours dans le but de les faire correspondre à un son cible fourni en entrée. La connaissance instrumentale est basée sur le spectre harmonique moyen calculé sur la partie entretenue des instruments. L'algorithme est basé sur une décomposition en valeurs singulières (SVD) de la contribution du spectre. La cible est alors exprimée comme une somme pondérée des spectres contenus dans la base de données. Un autre système a été proposé par Hummel [36]. Un algorithme itératif est également utilisé, mais la distance est basée sur l'enveloppe spectrale plutôt que sur les partiels. Le système calcule l'enveloppe spectrale de la cible et trouve itérativement la meilleure approximation. En conséquence, les structures harmoniques peuvent être très différentes entre l'entrée et le mélange obtenu. L'auteur conseille donc d'utiliser son système sur des sons dépourvus de hauteur tels que des voyelles chuchotées.

Récemment, une approche intéressante a été proposée par Carpentier [14]. Bien que la formulation du problème reste la même [12], le système est basé sur l'exploration génétique [66] ce qui permet de proposer un ensemble de solutions optimales, plutôt qu'un unique mélange instrumental. La cible est représentée par un ensemble de descripteurs (*features*) audios et symboliques qui décrivent les différents aspects du son à reproduire avec un orchestre prédéterminé. En utilisant une optimisation multi-critères [13], le système extrait le front de Pareto des solutions, qui représente la frontière des solutions non-dominées dans les diverses dimensions. L'utilisateur choisit alors une solution, qui permet au système d'en déduire ses préférences parmi les différentes dimensions. Le savoir instrumental est fondé sur le modèle développé par Tardieu dans [68]. Ce modèle utilise les *Gaussian Mixture Models* (GMM) [65] et permet d'apprendre la distribution des descripteurs pour un grand nombre de techniques de jeu. En outre, un ensemble de fonctions [67] ont été élaborées pour chaque descripteur et permettent de prédire les caractéristiques d'un mélange de sons instrumentaux.

2.4.3 Discussion

Nous tenterons ici de mettre en exergue les différentes faiblesses des systèmes d'orchestration existants.

- Il est à chaque fois question d'approcher une "cible sonore". C'est là se restreindre à un cas très particulier de l'orchestration, car la plupart du temps, le compositeur ne dispose pas au préalable d'un tel son. L'orchestration, en général, a pour but de produire, non de re-produire, un timbre. Le processus relatif à l'orchestration, s'il a un objectif précis, n'a pas de point de départ systématique et comme Claude Cadoz le disait, "*le processus de création est aussi celui de la genèse des idées. Un outil de création doit permettre cette genèse.*"
- Pour chacun de ces systèmes, la connaissance instrumentale se résume à un ensemble de sons instrumentaux. Sans même parler de leur accessibilité, la généricité de ces bases de données mérite d'être questionnée. Ne court-on pas alors un risque de "sur-apprentissage" ? Quelle est

la part variationnelle liée aux conditions d'enregistrement dans ce que le système "apprend" de l'échantillon ? Les orchestrations obtenues par ces outils produiront-elles les mêmes effets de timbre dans des conditions de jeu différentes ?

- Les trois premiers systèmes formulent l'hypothèse implicite de l'additivité des timbres instrumentaux. Or la simple considération de l'effet de phase suffit à la mettre en défaut. Les compositeurs connaissent bien ce phénomène. Il est donc permis de douter des capacités prédictives d'un modèle additif, aussi avantageux soit-il d'un point de vue mathématique.
- Les effets de salle (notamment la réverbération) ne sont pris en compte dans aucun de ces systèmes. Or, la réverbération naturelle des lieux de concert a un impact énorme sur les phénomènes de fusion acoustique. Les compositeurs le savent bien, et comptent les effets de salle parmi leurs alliés pour la réussite de leurs orchestrations.

Il est important de noter que tous les travaux antérieurs sur ce sujet ont mis l'accent sur une orchestration *verticale*, partant de l'analyse ciblée sur la partie entretenue des instruments pour trouver la meilleure combinaison se rapprochant de la cible, ignorant ainsi complètement l'évolution temporelle des sons. Tous les systèmes limitent ainsi la connaissance instrumentale aux sons harmoniques, entretenus et sans variations temporelles. Toutefois, Helmholtz [34], même limité par ses moyens d'investigation, avait déjà remarqué en 1863 que la reconnaissance d'un son était liée à son évolution temporelle plutôt qu'au seul spectre de sa partie stationnaire. En effet, la musique n'est pas seulement une collection de moments statiques, mais bien une évolution dans le temps. En outre, la notion de timbre en lui-même n'existe que dans la dimension du temps puisque les propriétés spectrales prennent naissance par l'évolution temporelle des ondes. Il convient également de noter que, dans sa pratique quotidienne de l'orchestration, le compositeur ne travaille pas uniquement par superposition verticale de sons stationnaires, mais bien plutôt par une sculpture simultanément verticale et horizontale du son. Pour toutes ces raisons, nous souhaitons aborder le problème du temps dans l'orchestration musicale. Cependant, nous devons garder à l'esprit que les phénomènes ayant une évolution dans le temps sont toujours restés difficiles à formaliser.

2.5 Notre approche

Comme nous l'avons vu tout au long de ce chapitre, l'orchestration se situe à la croisée de nombreuses disciplines scientifiques et fait naître de multiples axes de complexité. Si nous essayons de saisir tous les enjeux et questions soulevés par l'orchestration, nous pouvons rapidement nous apercevoir de l'immensité du champ d'étude abordé. Ainsi trois grands axes semblent sous-tendre la question de l'orchestration, le rapport entre signal et symbolique, le lien entre micro et macro-temps et enfin la composition horizontale à verticale du matériau sonore. Car cette discipline, au delà de sa formulation classique reposant sur les instruments de l'orchestre nous amène à étudier les mélanges sonores de toutes origines, leur organisation, la manière d'interagir avec eux, notre perception et appréciation de telles combinaisons ainsi que des phénomènes qui en émergent.

En complément de l'intérêt artistique de ce projet, l'orchestration se présente comme un cas d'étude idéal reliant diverses disciplines scientifiques. Chacune des disciplines concernées pourra ainsi tirer profit des questions sous-jacentes à l'orchestration. Par exemple, d'un point de vue psychoacoustique, l'étude des notions propres à l'orchestration, telles que la fusion, le volume, la diffraction, entre autres, reste inexistant de nos jours. Au niveau du traitement du signal, une connaissance accrue du timbre est nécessaire et appelle à la mise en place de nouveaux modèles d'évolution temporelle permettant de modéliser de manière précise la dynamique des instruments musicaux. Ainsi l'intégration de la composante micro-temporelle du timbre est une condition nécessaire pour envisager l'utilisation d'un outil informatique d'orchestration en situation de concert. D'une manière plus générale, l'étude des possibilités de mélanges sonores peut également être appliquée aux systèmes de synthèse sonore. Au niveau informatique, l'orchestration recoupe les domaines des bases de données, des interfaces de contrôle, de la programmation par contrainte et les diverses techniques appliquées aux domaines symbolique et du

traitement du signal devront coexister, et éventuellement donner naissance à des techniques hybrides inespérées. À titre d'exemple, le contrôle des systèmes issus de la recherche opérationnelle par des langages de programmation (e.g. par contraintes) nous fait croire à la création de nouveaux paradigmes expressifs combinant l'extraction sémantique automatique avec le contrôle pensé de la programmation. Au niveau de la symbolique, l'intégration de la composante macro-temporelle d'une forme visée permettra de mettre en rapport écriture et timbre, dans le cadre d'une cible orchestrale. La définition d'un système de raisonnement qualitatif composé d'un ensemble de termes pertinents pour l'orchestration (tessitures, zones de fusion, de diffraction, mise en relief, en volume) et l'amélioration de la connaissance sur le savoir instrumental, permettrait d'imaginer la mise en place d'un outil d'instrumentation moderne particulièrement recherché par les compositeurs. Enfin l'étude de la question séculaire de l'intention du compositeur dans la création musicale, par une concertation avec de nombreux artistes, permettrait de mettre en place de nouvelles interfaces de contrôle et de nouvelles techniques d'interaction musicale.

Dans un premier temps, nous conserverons pour la durée de ce stage l'approche qui a été prise par les anciens systèmes d'orchestration en considérant un son cible spécifié par l'utilisateur, représentant le timbre désiré. Le système devra alors trouver le mélange instrumental le plus approprié dans une base de données d'échantillons. Plusieurs bases de données nous permettent d'envisager l'utilisation d'un ensemble exhaustif d'instruments et de modes de jeux. Ainsi la base *Studio On Line* [4] (SOL) mise en place à l'IRCAM, la *Vienna Symphonic Library* [79] ainsi que la base *Real World Computing* [28] (RWC) permettent d'envisager un savoir instrumental presque complet. Nous utilisons ainsi la base de données mise en place pour le projet *Orchidée* que nous augmentons considérablement avec un ensemble de sons à variation temporelle fortement marquée.

Notre approche consiste alors à introduire la composante temporelle dans l'orchestration, condition nécessaire pour envisager l'orchestration de séquences polyphoniques. Deux composantes du temps peuvent alors se dégager. La composante *micro-temporelle* consiste en la modélisation de l'évolution temporelle d'un instrument. Ce modèle permettrait alors de dépasser la contrainte harmonique entretenue en introduisant des modes de jeu tels que le vibrato ou le glissando. La deuxième composante est *macro-temporelle* celle ci représente l'évolution complète d'une séquence au cours du temps. Toutefois, notre approche dans ce stage ne peut être vue que comme un système partiel, dans lequel l'utilisateur spécifie le son à reproduire (la cible) et les instruments qui devraient être utilisées pour trouver le meilleur mélange. Nous utilisons une base de données d'échantillons instrumentaux qui doit contenir autant de hauteurs, intensités et styles de jeu que possible pour chaque instrument. Nous présentons donc d'abord un modèle instrumental basé sur les mixtures gaussiennes permettant de représenter les propriétés à la fois spectrales et temporelles d'un instrument, puis nous comparons et validons celui-ci par rapport aux systèmes existants. Ce modèle permet en outre de capturer les propriétés pertinentes des timbres instrumentaux avec une grande réduction dimensionnelle tout en permettant une resynthèse efficace de ceux-ci. De plus, le point fort de ce modèle est de permettre d'inférer de manière quasi automatique la structure temporelle d'un son identique mais de durée modifiée tout en conservant les segments d'attaque et de relâchement. Nous présentons par la suite les résultats d'expériences issues de l'intégration de nos travaux dans le système d'orchestration existant *Orchidée* en notant l'amélioration apportée par rapport à l'orchestration "statique". Ainsi notre recherche s'engageant dans la voie du temps, il nous faut abandonner, ne serait-ce que temporairement, l'approche combinatoire, verticale, de l'orchestration pour se concentrer sur une modélisation du temps. En inversant le paradigme, l'orchestration musicale peut être considérée comme un problème de séparation de sources où le but est de trouver la combinaison d'instruments qui explique au mieux la cible. Compte tenu de cette approche, nous utilisons la factorisation en matrices non-négatives pour permettre de décomposer le spectre de la cible en une somme pondérée des instruments choisis. L'objectif est d'abord de découper le spectre en un ensemble d'unités temporelles qui ont une certaine cohérence spectrale. Cette approche permet à l'entrée cible d'être une chanson complète ayant plusieurs éléments à orchestrer séparément. De cette façon, notre système peut être utilisé dans un optique de réorchestration. Chaque unité est analysée avec la même transformation que celle utilisée pour les échantillons instrumentaux. Nous finissons par obtenir une représentation sous la forme d'une partition où chaque unité est orchestrée séparément et la durée de chaque élément reste la même que l'originale.

Chapter 3

Modèle micro-temporel

La musique est un processus articulé, ce qui inclus de nos jours non seulement la définition de structures macro-temporelles mais également la compréhension du domaine du micro-temps relatif aux qualités du timbre. Par conséquent, si nous pensons le timbre comme partie intégrante du processus compositionnel, nous devons nécessairement étendre à cette dimension celle de son évolution pour composer à des échelles temporelles supérieures. Comme nous le disions, le territoire du timbre n'est pas confiné à une structure statique des proportions, c'est à dire un arrangement fixe des fréquences de partiels, mais comprends des "lois de variation" qui régulent l'interaction entre les fonctions de fréquence et d'amplitude dans un contexte variable au cours du temps. Les modèles doivent être raffinés pour saisir le timbre en tant que structure spectro-temporelle.

3.1 Modèles existants

Lorsqu'on s'intéresse aux modèles instrumentaux existants, on se rend vite compte de l'absence presque complète de modélisation temporelle. Les modèles étant le plus souvent destinés à la discrimination où la classification [52], les instruments sont généralement représentés de manière simpliste avec un ensemble de descripteurs spectraux. Le modèle instrumental de Tardieu [68] utilisé par le système d'orchestration de Carpentier [14] omet par exemple complètement l'information temporelle. Les sons utilisés sont donc considérés stationnaires et sans évolution temporelle. Pour arriver à modéliser cette information, on peut alors imaginer le modèle simpliste ADSR [32] appliqué à chaque partiel, cependant malgré une paramétrisation avancée [39] ses performances en termes de précision restent très limitées. Une amélioration consistant à utiliser des polynômes de Legendre [70] permet d'augmenter la précision du modèle, et cette technique nous servira de point de comparaison avec les modèles proposés. Une autre méthode de modélisation temporelle consiste à calculer le spectre de modulation [64], déjà utilisé pour l'analyse multipitch [18] et la séparation de source [3]. Malheureusement celui-ci provoque une augmentation considérable des données nécessaires ce qui est en contradiction avec le but d'un modèle. D'autres approches peuvent être envisagées, comme par exemple un modèle source-filtre [76] qui permettrait de conserver l'information temporelle, tout en réduisant la dimensionnalité des données, cependant ce modèle limite la capacité de modélisation à certains modes de production. Il serait donc profitable d'utiliser des techniques de représentation telles que le modèle Harmonic Temporal Clustering (HTC) [41], qui permet de représenter l'information instrumentale sous forme d'une mixture de gaussiennes aussi bien d'un point de vue harmonique que temporel. Ce modèle ayant originellement été développé pour l'analyse multipitch [44] puis développé pour l'analyse de scènes auditives [49] présente de nombreuses propriétés intéressantes. Il permet notamment dans sa version originale de déterminer les attaques, hauteurs, durées et instruments présents dans un signal. Il a par ailleurs été étendu à l'analyse et la manipulation d'échantillons instrumentaux [1] et permet ainsi la synthèse de sons aux propriétés modifiées avec un minimum de distorsion.

3.2 L'incertitude d'un modèle

John Strawn [63] notait qu'aucun des algorithmes de réduction dimensionnelle ne peut *“prendre en compte des considérations à la fois globales et locales”* puisque *“une approximation utilisant un seuil très bas capture des nombreux détails”*, mais si *“le seuil est augmenté, des détails sont perdus mais la forme générale de la fonction est mise en valeur”*. Il est particulièrement difficile de *“trouver un paradigme de contrôle du seuil pour que les particularités nécessaires soient retenues mais que les détails superflus ne le soient pas.”* Nous avons donc ici à faire face à une sorte de “principe d'incertitude” analogue au fameux principe introduit pour la physique quantique par Heisenberg au début du siècle. Il est intéressant de noter que le même problème générique peut être rencontré dans n'importe quel type de représentation. Outre la dualité temps / fréquence, ce principe peut être transposé comme l'impossibilité de représenter au même moment des caractéristiques locales et globales. Si nous nous concentrons sur des caractéristiques locales, nous risquons alors de nous perdre dans l'information superflue (phénomène de “sur-apprentissage”) et la forme générale sera peu prise en compte. Si au contraire nous voulons représenter uniquement la forme générale, certaines des particularités pertinentes seront irrémédiablement perdues.

3.3 Diriger l'incertitude

Pourtant, face à ce problème d'incertitude, nous pouvons nous tourner vers les études perceptives qui permettraient d'orienter notre discrimination des particularités superflues. Ainsi par la vérification psychoacoustique, nous pouvons contourner cette incertitude en utilisant des hypothèses reposant sur des bases expérimentales solides. Une étude particulièrement pertinente dans le cadre de la modélisation instrumentale est celle réalisée par McAdams et Beauchamp [46] qui s'intéresse à la discrimination de sons instrumentaux aux paramètres spectro-temporels simplifiés. Cette étude nous permet par conséquent d'entrevoir l'importance perceptuelle relative des diverses particularités spectro-temporelles et donc quelles caractéristiques peuvent être simplifiées. Les auteurs proposent six simplifications de données effectuées sur les paramètres temporels et fréquentiels :

1. **Lissage des enveloppes d'amplitude (LA)** : Cette opération permet de supprimer les microvariations dans les enveloppes d'amplitude des différents partiels au cours du temps. Ce lissage est effectué grâce à un filtrage passe-bas des enveloppes.
2. **Cohérence des enveloppes d'amplitude (CA)** : L'objectif est ici de supprimer les différences entre les enveloppes sans modifier l'intensité du son. Les enveloppes des partiels sont donc toutes remplacées par une même fonction proportionnelle à l'enveloppe d'intensité et qui dépend de l'amplitude moyenne de chaque partiel.
3. **Lissage des enveloppes spectrales (LS)** : Cette opération vise à conserver les microvariations mais les enveloppes spectrales sont ici moyennées par rapport aux autres permettant d'obtenir une évolution conjointe moyenne des différentes enveloppes.
4. **Lissage des enveloppes fréquentielles (LF)** : L'objectif est de tester l'importance des microvariations de fréquence pour un partiel. Les variations de fréquence sont donc lissées indépendamment pour chaque partiel, de la même manière que la première simplification.
5. **Cohérence des enveloppes fréquentielles (CF)** : Cette simplification permet de tester la discriminabilité de l'inharmonicité parmi les partiels d'un son. Ainsi le son obtenu contient des partiels liés par une relation harmonique parfaite.
6. **Platitude des enveloppes fréquentielles (PF)** : Cette opération permet de combiner la suppression de l'inharmonicité et des microvariations fréquentielles.

Les simplifications ont été appliquées à différents instruments et il est ensuite demandé aux sujets de discriminer les sons simplifiés des sons originaux

Simplification	Instruments							Moyenne
	Cl	Fl	Hc	Mb	Ob	Tp	Vn	
LA	0.56	0.80	0.79	0.59	0.54	0.73	0.59	0.66
CA	0.81	0.96	0.97	0.97	0.75	0.98	0.95	0.91
LS	0.98	0.97	0.96	0.99	0.99	0.82	0.99	0.96
LF	0.69	0.72	0.83	0.50	0.53	0.77	0.70	0.67
CF	0.56	0.59	0.84	0.67	0.72	0.81	0.69	0.70
PF	0.70	0.72	0.91	0.62	0.48	0.82	0.73	0.71

Table 3.1: Résultats de discrimination pour différentes simplifications spectro-temporelles

De ces résultats nous pouvons tirer un certain nombre de directions pour la modélisation. Tout d’abord, deux simplifications semblent extrêmement discriminantes. D’un côté la cohérence des enveloppes d’amplitude (CA) et de l’autre le lissage des enveloppes spectrales (LS) indiquent que l’évolution indépendante de l’amplitude des partiels est une caractéristique fondamentale. Ces résultats nous montrent qu’il est crucial de traiter indépendamment l’enveloppe de chaque partiel. La cohérence des enveloppes d’amplitude nous montre également que l’approche se limitant aux amplitudes moyennes des partiels est largement insuffisante.

A l’opposé, les microvariations aussi bien au niveau de l’amplitude (LA) que de la fréquence (LF) semblent être beaucoup moins discriminantes perceptivement. Il est donc possible de lisser (jusqu’à un certain niveau) ces différentes enveloppes tout en conservant l’information adéquate. Enfin, même si les simplifications sur l’inharmonicité semblent également avoir un taux de discrimination faible, il faut tenir compte du peu d’inharmonicité naturelle présente dans les instruments testés. De plus, partant de l’hypothèse que les enveloppes des partiels doivent être traitées séparément, la réduction dimensionnelle engendrée par une telle simplification semble négligeable.

Nous étudierons donc la modélisation suivant les deux axes de simplification permettant une discrimination perceptive minimale, c’est à dire les microvariations fréquentielles et temporelles.

3.4 Microvariations fréquentielles

Il semble que notre sensibilité aux variations soit grandement dépendante de la fréquence à laquelle celles-ci se produisent. Cette sensibilité variable nous incite à considérer une échelle de fréquence musicale pour analyser les microvariations fréquentielles.

3.4.1 Transformée à Q constant

Les fréquences qui composent l’échelle de la musique occidentale sont géométriquement espacées. Ainsi même si la transformée de Fourier rapide (FFT) permet un calcul extrêmement rapide, les composantes obtenues ne correspondent pas efficacement aux fréquences musicales. Cette transformée donne pour chaque bande de fréquence une résolution constante égale à la fréquence d’échantillonnage divisée par la taille de la fenêtre en nombre d’échantillons. Cependant pour être adaptée au tempérament égal, la résolution devrait être géométriquement reliée à la fréquence. Cela signifie que le facteur entre fréquence et résolution $f/\delta f = Q$ soit constant. Un calcul similaire à la transformée de Fourier mais avec un facteur constant entre fréquence et résolution, nommé *transformée à Q constant* peut être obtenue. En utilisant cette transformée, il est possible d’obtenir une représentation temps-fréquence structurée à la dimensionnalité réduite. L’implémentation proposée par Brown [10] peut être vue comme un banc de filtres en $1/24^e$ d’octave. Ainsi, l’analyse permet aux composantes fréquentielles de correspondre aux quarts de tons dans l’échelle du tempérament égal et donc de lisser implicitement les microvariations fréquentielles. Dans ce banc de filtre à Q constant, la fréquence de la composante spectrale k_{cQ} est alors

$$\omega_{k_{cQ}} = \left(2^{1/24}\right)^{k_{cQ}} \omega_{min} \quad (3.1)$$

avec $\omega_{k_{cQ}}$ variant de la fréquence minimum choisie ω_{min} à une fréquence maximale inférieure à la fréquence de Nyquist. Dans la transformée de Fourier, la résolution ou bande passante δf est égale à la fréquence d'échantillonnage divisée par la taille de la fenêtre. Cette bande de fréquence dépend du choix de la fonction de fenêtrage $w[n, k_{cQ}]$ mais est inversement proportionnelle à la taille de la fenêtre $N[k_{cQ}]$. On peut ainsi conserver Q constant en choisissant les valeurs de $N[k_{cQ}]$ pour être inversement proportionnelles à celles de $\omega_{k_{cQ}}$. Pour une précision en quart de ton, on obtient

$$Q = \omega / \delta\omega = 1 / (2^{1/24} - 1) \approx 34 \quad (3.2)$$

Ainsi avec une fréquence d'échantillonnage S , la taille de la fenêtre pour la fréquence $\omega_{k_{cQ}}$ est donc

$$N[k_{cQ}] = S / \delta\omega_{k_{cQ}} = (S / \omega_{k_{cQ}}) Q \quad (3.3)$$

En utilisant l'approche par banc de filtres on obtient donc pour la composante spectrale k_{cQ}

$$X^{cQ}[k_{cQ}] = \sum_{n=0}^{N[k_{cQ}]-1} w[n, k_{cQ}] x[n] e^{-j\omega_{k_{cQ}} n} \quad (3.4)$$

l'exponentielle ayant ici l'effet d'un filtre centré sur la fréquence $\omega_{k_{cQ}}$. En prenant en compte les contraintes sur la fréquence et la taille de la fenêtre, la fréquence de la composante spectrale k_{cQ} est $2\pi Q / N[k_{cQ}]$. La fonction de fenêtrage est de la même forme pour toutes les composantes mais sa durée est déterminée par $N[k]$ qui est une fonction à la fois de k et de n . On obtient alors après normalisation

$$X^{cQ}[k_{cQ}] = \frac{1}{N[k_{cQ}]} \sum_{n=0}^{N[k]-1} w[n, k_{cQ}] x[n] e^{-j\omega_{k_{cQ}} n} \quad (3.5)$$

Cette équation est utilisée comme base pour le calcul avec $N[k_{cQ}] = N_{max} / (2^{1/24})^{k_{cQ}}$. La condition de Nyquist devient alors $2\pi Q / N[k_{cQ}] < \pi$, ce qui signifie que $N[k_{cQ}] > 2Q$. On utilise une fenêtre de Hamming qui est définie par

$$w[n, k_{cQ}] = \alpha + (1 - \alpha) \cos(2\pi n / N[k_{cQ}]) \quad (3.6)$$

avec $\alpha = 25/46$ et $0 \leq n \leq N[k] - 1$. La méthode de calcul par banc de filtres est cependant très coûteuse en calcul. Une méthode plus efficace a été introduite par Brown et Puckette [11]. Celle-ci implique le calcul d'un ensemble de noyaux qui sont ensuite appliqués à chaque FFT. Par conséquent le temps de calcul total n'est que légèrement plus long que celui d'une transformée de Fourier car seules quelques multiplications sont impliquées dans le calcul de chaque composante. En utilisant une variante de l'équation de Parseval, il peut être montré que pour n'importe quelle fonction discrète $x[n]$, $y[n]$ et sa transformée de Fourier $X[k]$ et $Y[k]$, on a

$$\sum_{n=0}^{N-1} x[n] y^*[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] Y^*[k] \quad (3.7)$$

En utilisant cette équation et en posant

$$w[n, k_{cQ}] e^{-j\omega_{k_{cQ}} n} = \mathcal{K}^*[n, k_{cQ}] \quad (3.8)$$

On peut ainsi obtenir l'équation des noyaux $K[k, k_{cQ}]$ dans le domaine spectral :

$$K[k, k_{cQ}] = \sum_{n=0}^{N-1} w[n, k_{cQ}] e^{j\omega_{k_{cQ}} n} e^{-j2\pi kn / N} \quad (3.9)$$

Les noyaux sont normalisés par la longueur de la fenêtre pour que les coefficients sinusoïdaux dans le signal aient la même amplitude dans la transformée à Q constant obtenue. Les résultats d'une telle

transformée et sa comparaison avec la transformée de Fourier sont disponibles en annexe. Nous utilisons pour notre modèle une précision d'un seizième de ton (96 bandes par octave) pour des fréquences allant de $\omega_{min} = 60Hz$ à $\omega_{max} = 8000Hz$.

Grâce à cette transformée, nous lisons donc implicitement les micro-variations fréquentielles et réduisons également grandement la dimensionnalité des données. Cependant, la dimensionnalité reste encore importante, ceci est dû à la conservation complète de l'information temporelle.

3.5 Microvariations temporelles

Nous discutons ici de différentes modélisations possibles pour l'enveloppe d'amplitude des partiels. Dans le cas général où le type de source n'est pas connu à l'avance, il est sûrement plus sage de ne pas utiliser un modèle physique basé sur un mécanisme de production sonore particulier. Il est donc important d'introduire un modèle aussi générique que possible.

3.5.1 Approximation polynomiale

Une première approximation peut être faite en utilisant un modèle polynomial. L'approximation polynomiale consiste à trouver les coefficients d'un polynôme $p(x)$ de degré n qui correspond aux données y au sens des moindres carrés. Le modèle obtenu est alors un vecteur de taille $n + 1$ contenant les coefficients polynomiaux p_k tels que

$$p(x) = p_1x^n + p_2x^{n-1} + \dots + p_nx + p_{n+1} \quad (3.10)$$

Pour ce faire, il est possible de former la matrice de Vandermonde V dont les éléments sont les puissances de x soit

$$v_{i,j} = x_i^{n-j} \quad (3.11)$$

Puis on résout alors le problème des moindres carrés

$$Vp \approx y \quad (3.12)$$

Les polynômes de Legendre $P_n(x)$ forment une base complète de fonctions orthogonales sur l'intervalle $-1 < x < 1$. Ces polynômes sont les solutions de l'équation différentielle de Legendre

$$\frac{d}{dx} \left[(1-x^2) \frac{d}{dx} P_n(x) \right] + n(n+1) P_n(x) = 0 \quad (3.13)$$

Nous utilisons donc les polynômes de Legendre pour modéliser de manière rapide l'évolution des partiels et avec un minimum de paramètres. Ce modèle nous servira également de référence pour la comparaison des performances vis-à-vis des autres modèles proposés.

3.5.2 Modèles de mixtures gaussiennes

3.5.2.1 Distribution gaussienne

En mathématiques, la fonction gaussienne est bien connue dans le domaine des probabilités. Sa densité de probabilité à une dimension est définie avec la moyenne μ et la variance σ^2 par

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (3.14)$$

La distribution gaussienne est généralement une bonne approximation de la forme d'une classe de fonctions caractéristiques. Il s'agit d'une fonction mathématique solide qui s'étend facilement à de multiples dimensions. Cependant la distribution gaussienne est sous-tendue par l'hypothèse que la classe de fonctions à modéliser est simpliste. Si le modèle à obtenir est multimodal alors celle-ci échoue.

3.5.2.2 Modèles de mixtures gaussiennes

Les modèles de mixtures gaussiennes (GMM) sont un mélange de plusieurs distributions gaussiennes et peuvent donc représenter différentes sous-classes plus complexes. La fonction de densité de probabilité est alors définie comme une somme pondérée de gaussiennes

$$p(x; \theta) = \sum_{c=1}^C \alpha_c \mathcal{N}(x; \mu_c, \Sigma_c) \quad (3.15)$$

avec C le nombre de composantes, α_c le poids de la composante c , $0 < \alpha_c < 1$ pour toutes les composantes et $\sum_{c=1}^C \alpha_c = 1$. La liste de paramètres

$$\theta = \{\alpha_1, \mu_1, \Sigma_1, \dots, \alpha_C, \mu_C, \Sigma_C\} \quad (3.16)$$

définit un modèle de mixture gaussienne particulier. L'estimation de ces paramètres peut être considérée comme un apprentissage non-supervisé dans le cas où les échantillons sont générés par des composantes individuelles d'une distribution mélangée et sans connaissance de quel échantillon a été généré par quelle composante. Les techniques de clustering essayent généralement d'identifier les composantes exactes, mais les mixtures de gaussiennes peuvent également être utilisées pour approximer une distribution arbitraire.

3.5.2.3 Estimation du maximum de vraisemblance

Lors de l'estimation des paramètres, la sélection du modèle initial peut se faire par exemple en échantillonnant directement la valeur des données, mais l'ajustement des paramètres du modèle nécessite une certaine mesure de similarité, c'est-à-dire, à quel point la distribution correspond aux données observées. La vraisemblance des données est une telle mesure. Supposons qu'il existe un ensemble d'échantillons indépendants $X = \{x_1, \dots, x_N\}$ tirés à partir d'une seule distribution décrite par une fonction de densité de probabilité $p(x, \theta)$ où θ est la liste des paramètres du modèle. La fonction de vraisemblance

$$\mathcal{L}(X; \theta) = \prod_{n=1}^N p(x_n; \theta) \quad (3.17)$$

indique la vraisemblance d'observer les données X compte tenu de la distribution de X , ou plus précisément, étant donné la distribution des paramètres θ . Le but est donc de trouver $\hat{\theta}$ qui maximise la vraisemblance

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \mathcal{L}(X; \theta) \quad (3.18)$$

Habituellement cette fonction n'est pas maximisée directement mais à travers son logarithme

$$L(X; \theta) = \ln \mathcal{L}(X; \theta) = \sum_{n=1}^N \ln p(x_n; \theta) \quad (3.19)$$

qu'on appelle la fonction de log-vraisemblance et qui est analytiquement plus facile à manipuler. En raison de la monotonie de la fonction logarithme, la solution de l'équation de maximisation est la même en utilisant $L(X; \theta)$ ou $\mathcal{L}(X; \theta)$ [69].

Selon la probabilité $p(x; \theta)$, il pourrait être possible de trouver le maximum de manière analytique en mettant les dérivés de la fonction log-vraisemblance à zéro puis en effectuant la résolution de θ . Cela peut être aisément réalisé pour une unique gaussienne, ce qui mène à l'estimation intuitive de la moyenne et de la variance [20], mais généralement l'approche analytique est insoluble. Dans la pratique, une méthode itérative comme l'algorithme Expectation Maximization (EM) est utilisée. Cependant la maximisation de la probabilité peut, dans certains cas, conduire à des estimations singulières, ce qui est le problème fondamental des méthodes de maximisation de vraisemblance avec les modèles de mélanges gaussiens [22].

3.5.2.4 Estimation par Expectation Maximisation

L'algorithme d'Expectation Maximization (EM) est une méthode itérative pour l'estimation des paramètres du maximum de vraisemblance à partir de données incomplètes. Il peut également être utilisé pour traiter les cas où une approche analytique pour l'estimation du maximum de vraisemblance est infaisable, comme les mélanges gaussiens avec des inconnues et des matrices de covariance et de moyennes non restreintes.

Supposons que chaque échantillon d'apprentissage contienne des caractéristiques connues et d'autres absentes ou inconnues. Toutes les caractéristiques connues sont notées X et toutes les caractéristiques inconnues sont notées Y . L'étape d'expectation (E-step) de l'algorithme EM est de former la fonction

$$Q(\theta; \theta^i) \equiv E_Y [\ln \mathcal{L}(X, Y; \theta) \mid X; \theta^i] \quad (3.20)$$

où θ^i est l'estimation à l'itération précédente des paramètres de la distribution et θ est la variable pour une nouvelle estimation décrivant la distribution complète. $\mathcal{L}$ est la fonction de vraisemblance. Cette fonction calcule la vraisemblance des données en incluant les caractéristiques inconnues Y marginalisées par rapport à l'estimation courante de la distribution décrite par θ^i . L'étape de maximisation (M-step) cherche à maximiser $Q(\theta; \theta^i)$ par rapport à θ et donc de sélectionner

$$\theta^i = \underset{\theta}{\operatorname{argmax}} Q(\theta; \theta^i) \quad (3.21)$$

Ces étapes sont répétées jusqu'à ce qu'un critère de convergence soit atteint. Pour la critère de convergence, il est suggéré dans [20] de choisir un seuil approprié T et d'arrêter l'algorithme lorsque

$$Q(\theta^{i+1}; \theta^i) - Q(\theta^i; \theta^{i-1}) \leq T \quad (3.22)$$

Un autre critère est proposé dans [69] tel que

$$\|\theta^{i+1} - \theta^i\| \leq \epsilon \quad (3.23)$$

pour un choix d'une norme et d'un ϵ adéquats. Le point commun à ces deux critères est que les itérations sont interrompues lorsque l'évolution des différentes valeurs tombe en dessous d'un seuil. Un critère plus sophistiqué peut être dérivé en utilisant un taux de changement relatif plutôt qu'absolu.

L'algorithme EM démarre à partir d'une estimation initiale de θ^0 pour la distribution des paramètres et la log-vraisemblance est garantie d'augmenter à chaque itération jusqu'à convergence. Cette convergence conduit à un maximum local ou global, mais peut aussi conduire à des estimations singulières, ce qui est particulièrement vrai pour les distributions de mélanges gaussiens arbitraires avec des matrices de covariance arbitraires. La description détaillée de l'algorithme EM général et de son application aux modèles de mélanges gaussiens peuvent être trouvés dans [20, 22, 69].

L'initialisation est l'un des problèmes majeurs de l'algorithme EM. Le choix de θ^0 détermine (en partie) si l'algorithme converge ou s'il atteint les limites de l'espace des paramètres, produisant ainsi des résultats singuliers, sans aucune pertinence. Certaines solutions utilisent de multiples points de dépars aléatoires ou commencent l'algorithme par le résultat d'un clustering.

3.5.2.5 Variance libre

L'application de l'algorithme EM aux mélanges gaussiens [24] se déroule de la manière suivante. Les données connues X sont interprétées comme des données incomplètes. La partie manquante Y est la connaissance de quel élément produit chaque échantillon x_n . Pour chaque x_n existe un vecteur binaire $y_n = \{y_{n,1}, \dots, y_{n,C}\}$, avec $y_{n,c} = 1$ si l'échantillon a été produit par la composante c , ou zéro autrement. La log-vraisemblance sur l'ensemble des données est alors

$$\ln \mathcal{L}(X, Y; \theta) = \sum_{n=1}^N \sum_{c=1}^C y_{n,c} \ln(\alpha_c p(x_n \mid c; \theta)) \quad (3.24)$$

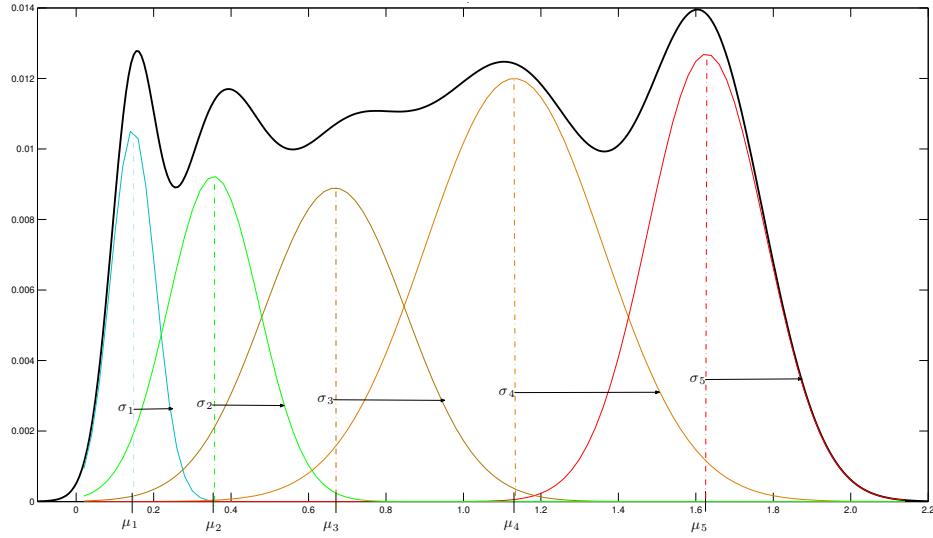


Figure 3.1: Modélisation d'une enveloppe par mélange gaussien à variance libre

L'étape expectation consiste à calculer l'espérance conditionnelle de la log-vraisemblance de l'ensemble des données, la fonction Q , étant donné X et l'estimation actuelle θ^i des paramètres. Étant donné que la log-vraisemblance complète $\ln \mathcal{L}(X, Y; \theta)$ est linéaire par rapport aux données manquantes Y , l'espérance conditionnelle $W \equiv E[Y | X, \theta]$ doit tout simplement être calculée et remplacée dans $\ln \mathcal{L}(X, Y; \theta)$. Par conséquent

$$Q(\theta, \theta^i) \equiv E[\ln \mathcal{L}(X, Y; \theta) | X, \theta^i] = \ln \mathcal{L}(X, W; \theta) \quad (3.25)$$

où les éléments de W sont définis par

$$w_{n,c} \equiv E[y_{n,c} | X, \theta^i] = \Pr[y_{n,c} = 1 | x_n, \theta^i] \quad (3.26)$$

La probabilité peut être calculée grâce à la loi de Bayes

$$w_{n,c} = \frac{\alpha_c^i p(x_n | c; \theta^i)}{\sum_{j=1}^C \alpha_j^i p(x_n | j; \theta^i)} \quad (3.27)$$

où α_c^i est la probabilité a priori (de l'estimation θ^i) et $w_{n,c}$ est la probabilité a posteriori que $y_{n,c} = 1$ après observation de x_n . En d'autres termes, $w_{n,c}$ est la probabilité que x_n soit produit par la composante c . En appliquant l'étape de maximisation pour l'estimation de la distribution des paramètres pour un mélange de C composantes gaussiennes de matrices de covariance arbitraires, les formules résultantes d'itération sont les suivantes :

$$\alpha_c^{i+1} = \frac{1}{N} \sum_{n=1}^N w_{n,c} \quad (3.28)$$

$$\mu_c^{i+1} = \frac{\sum_{n=1}^N x_n w_{n,c}}{\sum_{n=1}^N w_{n,c}} \quad (3.29)$$

$$\Sigma_c^{i+1} = \frac{\sum_{n=1}^N w_{n,c} (x_n - \mu_c^{i+1}) (x_n - \mu_c^{i+1})^T}{\sum_{n=1}^N w_{n,c}} \quad (3.30)$$

Les nouvelles estimations sont réunies dans θ^i . Si le critère de convergence n'est pas satisfait alors $i \leftarrow i + 1$ et les équations sont évaluées à nouveau avec les nouvelles estimations.

L'interprétation de ces équations peut être faite de la manière suivante. Le poids α_c d'une composante est la portion des échantillons appartenant à cette composante. Il est calculé par l'approximation de la probabilité conditionnelle pour chaque composante à la précédente estimation de paramètres et en prenant la probabilité postérieure de chaque échantillon appartenant à la composante c . Les échantillons sont pondérés avec leurs probabilités d'appartenance à la composante, puis la moyenne μ_c et la matrice de covariance Σ_c sont calculées.

Il est intéressant de noter que jusqu'à maintenant le nombre de composantes C était supposé connu. Pour remédier à ce problème, nous utilisons l'algorithme 'greedy EM' qui commence avec une composante unique, puis ajoute des composantes dans le mélange, une par une. La composante de départ optimale (au sens de la vraisemblance la plus élevée) est trivialement calculée. L'algorithme répète alors deux étapes : insérer une composante dans le mélange puis exécuter l'algorithme jusqu'à convergence. L'insertion d'une composante qui augmente au plus la vraisemblance est un problème bien plus facile que l'initialisation d'un ensemble proche de la répartition optimale. L'insertion de composante implique la recherche des paramètres pour une seule composante à la fois. Rappelons que l'algorithme EM trouve un optimum local pour les paramètres de la distribution, et pas nécessairement l'optimum global ce qui rend cette méthode sensible à l'initialisation.

Soit p_C un mélange de C composantes avec les paramètres θ_C . L'algorithme 'greedy EM' est le suivant [73]

1. Calculer le mélange optimal p_1 (au sens du maximum de vraisemblance) à une seule composante et mettre $C \leftarrow 1$.
2. Trouver une nouvelle composante $\mathcal{N}(\mathbf{x}; \mu', \Sigma')$ et la pondération correspondante α' qui accroît au maximum la vraisemblance :

$$\{\mu', \Sigma', \alpha'\} = \arg \max_{\{\mu, \Sigma, \alpha\}} \sum_{n=1}^N \ln [(1 - \alpha) p_C(x_n) + \alpha \mathcal{N}(\mathbf{x}_n; \mu, \Sigma)] \quad (3.31)$$

tout en gardant p_C fixe.

3. Affecter $p_{C+1}(x) \leftarrow (1 - \alpha') p_C(x) + \alpha' \mathcal{N}(\mathbf{x}; \mu', \Sigma')$ et $C \leftarrow C + 1$
4. Mettre à jour le mélange p_C à l'aide de l'algorithme EM jusqu'à convergence.
5. Évaluer un critère d'arrêt, passer à l'étape 2 ou quitter.

Le critère d'arrêt à l'étape 5 peut être par exemple n'importe quel modèle de sélection, nombre maximum de composantes ou encore vraisemblance maximale.

Le point crucial est bien sûr l'étape 2. Trouver la meilleure nouvelle composante nécessite une recherche globale, ce qui est effectué par la création de $C N_{cand}$ candidats. Le nombre de candidats augmente donc linéairement avec le nombre de composantes C , avec N_{cand} candidats pour chaque composante existante. Le candidat qui induit la plus grande vraisemblance lorsqu'il est inséré dans le mélange précédent est sélectionné. Les paramètres et le poids du meilleur candidat sont ensuite utilisés à l'étape 3 au lieu de la véritable valeur optimale [73].

Les candidats pour l'exécution de l'étape 2 sont initialisés comme suit : l'ensemble de données X est partitionné en C ensembles disjoint de données $\{A_c\}$, $c = 1 \dots C$, en fonction des probabilités a posteriori des composantes individuelles. Pour chaque A_c , N_{cand} candidats sont initialisés en choisissant de manière aléatoire deux points x_l et x_r dans A_c . L'ensemble A_c est alors divisé en deux en utilisant la plus petite distance par rapport à x_l et x_r . La moyenne et la covariance de ces deux nouveaux sous-ensembles sont les paramètres des deux nouveaux candidats. Les poids des candidats sont alors fixés à la moitié du poids de la composante qui produit l'ensemble A_c . Les nouveaux x_l et x_r sont tirés jusqu'à ce que N_{cand} candidats soient initialisés pour A_c . L'algorithme EM est alors utilisé sur chacun des candidats du nouvel ensemble. Afin de réduire la complexité de calcul de l'algorithme, une borne inférieure à la log-vraisemblance est utilisée à la place de la vraie log-vraisemblance. La limite inférieure de la

log-vraisemblance est calculée avec les points de chaque ensemble A_c . Les itérations sont interrompues lorsque la variation relative de log-vraisemblance du mélange à $C + 1$ composantes résultant tombe sous le seuil défini ou que le nombre maximal d'itérations est atteint. Lorsque l'algorithme EM a convergé, le candidat est prêt à être évalué [73]. Le nombre de composantes maximal est fixé à 10 pour éviter à la fois une solution singulière ainsi qu'un nombre de paramètres trop important.

3.5.2.6 Variance fixe

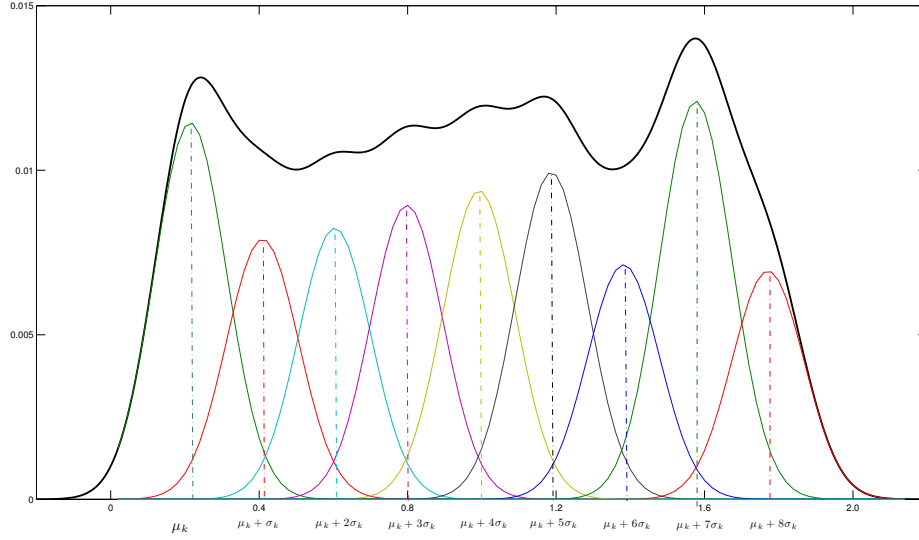


Figure 3.2: Modélisation d'une enveloppe par mélange gaussien à variance fixe

La particularité de ce modèle est que les centres des fonctions gaussiennes sont espacés d'une distance proportionnelle au paramètre de diffusion commun σ_k qui représente également l'écart-type de toutes les gaussiennes. On obtient alors le mélange suivant

$$U_k(x) = \sum_{c=1}^C \alpha_c \mathcal{N}(x; \mu_k + c \cdot \sigma_k, \sigma_k) \quad (3.32)$$

Cette liaison permet d'assurer la souplesse de la courbe en empêchant les gaussiennes d'être séparées les unes des autres. De plus, le modèle précédent nécessite de conserver trois paramètres par gaussienne utilisée, tout en limitant le nombre de composantes utilisables. Dans ce nouveau modèle, il n'est nécessaire de conserver que la moyenne de la première gaussienne μ_k ainsi que l'écart-type commun σ_k . De plus en interdisant les inversions de position des gaussiennes, cela réduit la singularité du système, améliorant ainsi les perspectives de modélisation. Enfin comme nous le verrons dans la section suivante, il devient facile de réaliser une modification de durée sur le modèle en utilisant le paramètre de diffusion σ_k . Nous utilisons donc le même algorithme EM que présenté précédemment mais en utilisant des formules de mise à jour modifiées de la manière suivante

$$\mu_k^i = \frac{\sum_{c=1}^C \sum_{n=1}^N (x_n - c \cdot \sigma_k^{i-1}) w_{n,c}}{\sum_{c=1}^C \sum_{n=1}^N w_{n,c}} \quad (3.33)$$

$$\sigma_k^i = \frac{1}{2} \left(\left((a_k^i)^2 + 4b_k^i \right)^{1/2} - a_k^i \right) \quad (3.34)$$

$$\begin{cases} a_k^i = \sum_{c=1}^C \sum_{n=1}^N c (x_n - \mu_k^i) w_{n,c} \\ b_k^i = \sum_{c=1}^C \sum_{n=1}^N (x_n - \mu_k^i)^2 w_{n,c} \end{cases}$$

L'équation de mise à jour des poids α_c des composantes reste inchangée. Il est intéressant de noter que si les paramètres de diffusion σ_k et de première moyenne μ_k sont constants entre les différents échantillons, on obtient alors un modèle parfaitement linéaire.

3.5.3 Modification de durée

Les modèles micro-temporels étant établis dans le but de comparer l'évolution spectro-temporelle, une question évidente apparaît. La majeure partie des instruments de l'orchestre étant à son entretien, il semble obligatoire que la comparaison entre la cible et le mélange instrumental soit effectuée sur une durée identique. En effet, si nous nous limitons à la durée originale des échantillons, l'indice de similarité serait inévitablement faussé par ces différences de durée. C'est ici qu'est révélé tout l'intérêt de travailler avec des mélanges de gaussiennes. En effet à partir des données du modèle obtenu, il devient facile d'inférer la structure d'un son instrumental identique mais de durée différente.

Nous considérons la modification de durée par expansion ou compression de l'enveloppe temporelle sur la partie entretenue des sons. Ainsi les segments d'attaque et de relâchement sont conservés en calculant leurs emplacements respectifs et les modifications ne s'appliquent qu'aux paramètres appartenant au segment entretenu.

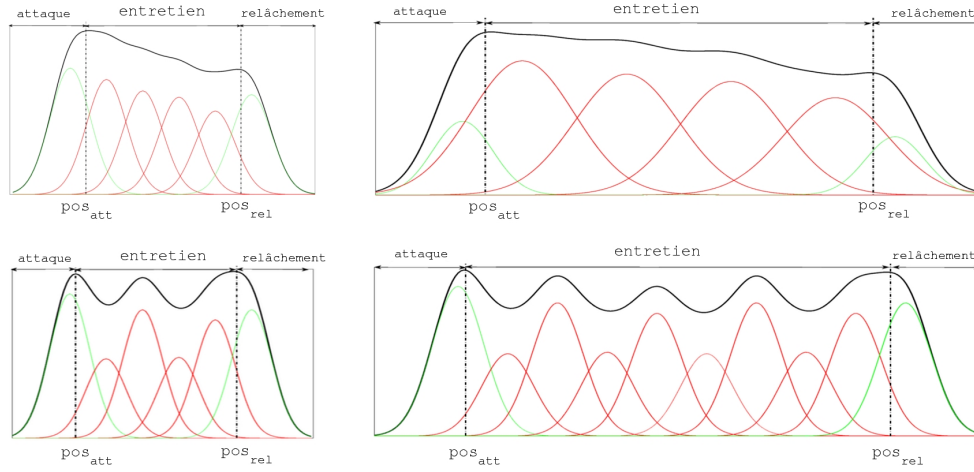


Figure 3.3: Deux approches de la modification de durée

Ce choix provient du fait que l'excitation durant les segments d'attaque et de relâchement sont généralement identiques indépendamment de la durée d'une note. Les gaussiennes modifiées sont alors celles dont la moyenne satisfait

$$pos_{att} + \epsilon < \mu_k < pos_{rel} - \epsilon$$

avec pos_{att} la position de la fin du segment d'attaque et pos_{rel} la position du début du segment de relâchement. On obtient l'ensemble $k \in K = \{\mathcal{N}(\mu_k, \sigma) / pos_{att} + \epsilon < \mu_k < pos_{rel} - \epsilon\}$. Deux méthodes différentes peuvent alors être appliquées pour la modification de durée. Comme nous le disions pour le modèle à variance fixe, le paramètre de diffusion σ_k peut être utilisé de manière simpliste pour effectuer une dilatation ou une compression des gaussiennes concernées. Si nous voulons modifier la durée d'un facteur γ , on obtient alors pour les gaussiennes de l'ensemble K

$$\sigma'_k = \gamma \sigma_k$$

Il faut alors modifier les poids des gaussiennes modifiées de l'ensemble K de la même manière pour compenser la modification de variance

$$\alpha'_k = \gamma \alpha_k$$

Si on se place dans le modèle à variance fixe, les moyennes des gaussiennes étant liées par le facteur de diffusion commun, il suffit alors de calculer pour $n \geq \min(k)$

$$\mu_n = \mu_{n-1} + 2\sigma'_{n-1}$$

Dans le cas du modèle à variance libre, il faut cette fois-ci recalculer chaque moyenne indépendamment en fonction du facteur de dilatation γ

$$\mu'_n = \gamma \cdot \mu_n + pos_{att} (1 - \gamma)$$

D'un autre côté il est possible pour les cas particuliers que sont le vibrato et le tremolo de conserver les paramètres originaux mais de modifier le nombre de gaussiennes et de mettre à jour les poids des composantes correspondantes. Il est cette fois nécessaire de modifier le nombre de gaussiennes présentes, nous devons donc calculer le nouveau nombre de composantes

$$N_{new} = \lceil \gamma \cdot N_k \rceil + N_{g \notin K}$$

Les paramètres des composantes pré-existantes restent alors inchangés et nous considérons que les poids des composantes suivent un comportement cyclique, soit

$$\alpha_n = \alpha_{n-card(K)}$$

Il est nécessaire de remarquer que ces deux approches s'appliquent chacune à des cas spécifiques. En effet, dans le cas d'une enveloppe linéaire, la modification de durée par répétition n'a aucune pertinence. Inversement, dans le cas du tremolo ou du vibrato, même si la modification par dilatation semble pouvoir inférer une structure "réalisable", il est évident que l'approche par répétition est beaucoup plus pertinente. Il semble donc que pour inférer correctement, le choix de la méthode de modification de durée devrait être relatif à une taxinomie des instruments en cours d'analyse. Une comparaison de la pertinence de ces deux approches sur des enveloppes de natures variées est disponible en annexe.

3.6 Résultats

3.6.1 Base de données

Pour tester les différents modèles présentés précédemment, il faut se placer d'emblée dans une situation d'importance équivalente à l'immensité orchestrale. Pour cela nous analysons donc une large base de données, issue des travaux de Carpentier et Tardieu [12] pour le système *Orchidée*, qui est décrite en détail dans le tableau suivant.

Type d'instrument	Modes de jeux	Nombre d'échantillons
Bassons	8	458
Clarinettes	18	814
Flutes	20	1.191
Cors	11	616
Hautbois	11	695
Saxophones	2	161
Contrebasses	21	2.498
Altos	25	3.017
Violons	26	2.911
Violoncelles	25	2.930
Trombones	20	1.187
Trompettes	18	795
Tubas	9	891
Total	215	18.164

Table 3.2: Base de données utilisée pour l'analyse

3.6.1.1 Extension de la base de données

La base de données originale du système *Orchidée* ne contient malheureusement aucun mode de jeu à évolution temporelle suffisamment marquée. Il nous a donc fallu étendre cette base avec de nouveaux échantillons représentant des modes de jeu particulièrement évolutif. Nous avons donc ajouté pour chaque instrument (si existant) les modes de vibrato, tremolo, glissando, trille, crescendo, decrescendo ainsi que des échantillons passant d'un mode de jeu à un autre.

Type d'instrument	Modes de jeux	Nombre d'échantillons
Bassons	3	68
Clarinettes	5	111
Hautbois	8	154
Saxophones	16	282
Contrebasses	17	415
Harpe	9	144
Altos	16	408
Violons	15	369
Violoncelles	16	304
Trombones	13	260
Trompettes	14	254
Tubas	9	144
Total	141	3.217

Table 3.3: Base de données à forte variation temporelle

3.6.2 Comparaison

Nous discutons ici les résultats des différents modèles proposés. Comme nous le disions précédemment, la mise en place d'un modèle passe toujours par de nombreux compromis. Il est cependant possible d'observer leurs performances sous différents angles. Tout d'abord nous insistons sur l'importance de la précision d'un modèle en calculant les divergences entre le son d'origine et le son resynthétisé à partir des données réduites. Cette précision doit être confrontée à la réduction dimensionnelle engendrée par le modèle qui permet d'évaluer la performance du modèle à capturer les caractéristiques essentielles. Enfin, nous comparons les performances en termes de temps de calcul, qui n'est pas un critère de choix central étant donné que l'analyse de la base de données n'est effectuée qu'une seule fois.

3.6.2.1 Précision

Pour vérifier la précision, nous commençons par resynthétiser le spectre grâce aux paramètres du modèle. La mesure de précision est effectuée à l'aide de deux mesures d'erreur différente, la norme L^2 et la divergence de *Kullback-Leibler* généralisée. Nous calculons ces deux mesures entre la transformée d'origine et la transformée resynthétisée. Cependant la durée et le contenu des échantillons pouvant être très disparates, nous normalisons ces mesures avec la divergence entre chaque échantillon et un silence de durée équivalente. De cette manière nous pouvons comparer les résultats avec des sons de nature différente. Enfin nous confrontons ces résultats avec le temps de resynthèse et la réduction dimensionnelle engendrée par le modèle. Le tableau suivant synthétise les résultats des analyses sur l'ensemble des 21.381 échantillons de la base de données.

Comme nous l'espérons, le modèle à variance fixe est celui qui nous permet d'obtenir la plus grande précision et ce avec une réduction dimensionnelle comparable, la précision du modèle à variance libre pourrait être augmentée en autorisant un nombre de composantes supérieur mais on augmente proportionnellement les possibilités d'obtenir des estimations singulières. D'autre part même si nous pourrions argumenter que l'ordre des polynômes de Legendre pourrait être augmenté, cela engendrerait des phénomènes d'oscillations qui fausseraient encore plus le résultat.

		Polynômes de Legendre	Gaussian Mixture Models	
			Variance libre	Variance fixe
Temps de resynthèse		0.0031 (+/- 0.0037)	0.0212 (+/- 0.0032)	0.0250 (+/- 0.0138)
Précision (divergence)	L^2	39.749 (+/- 836.91)	19.429 (+/- 218.31)	8.3071 (+/- 37.341)
	$L^2/Silence$	0.5690 (+/- 0.1529)	0.2143 (+/- 0.0267)	0.0897 (+/- 0.0223)
	KL	335.79 (+/- 6987.8)	272.82 (+/- 458.97)	201.72 (+/- 362.76)
	$KL/Silence$	0.9650 (+/- 0.5047)	0.6176 (+/- 0.0043)	0.3811 (+/- 0.0154)
Réduction dimensionnelle		77.236 (+/- 498.75)	52.202 (+/- 339.28)	48.687 (+/- 246.69)

Table 3.4: Temps de resynthèse, réduction dimensionnelle et précision des modèles

Les figures et statistiques complètes de comparaison sont disponibles en annexe.

3.6.2.2 Complexité calculatoire

Nous comparons ici les temps de calcul des différents modèles. Dans un premier temps, nous pouvons comparer les temps de calcul sur l'ensemble de la base de données. Les calculs ont été effectués sur un Mac OSX 10.5.6 avec un processeur 2.4 GHz Intel Core 2 Duo et 2 Go de mémoire DDR3.

	Bois	Cordes	Cuivres	Total
Legendre	2h 13m 13s	6h 16m 37s	1h 46m 57s	10h 16m 29s
GMM Fixe	20h 54m 39s	75h 30m 5s	14h 24m 53s	110h 43m 51s
GMM Libre	63h 12m 51s	234h 47m 23s	53h 20m 42s	351h 11m 20s

Table 3.5: Temps de calcul totaux par famille

Pour entrevoir la durée requise pour le calcul des différents modèles lors de l'ajout d'un instrument ou d'un mode de jeu, nous pouvons comparer les temps de calcul moyens pour une bande fréquentielle, un échantillon et un mode de jeu entier (considéré contenant 64 notes).

	Bande Fréquentielle	Echantillon	Mode de jeu
Legendre	0.062 (+/- 0.0024)	2.064 (+/- 0.3204)	577.922 (+/- 31.004)
GMM Fixe	0.631 (+/- 0.1169)	25.996 (+/- 27.9786)	6187.8 (+/- 275.13)
GMM Libre	1.5536 (+/- 0.2348)	83.5473 (+/- 297.15)	9677.0 (+/- 944.38)

Table 3.6: Temps de calcul moyens et variances en secondes

Comme nous le voyons sur ce tableau, l'approximation polynômiale est de loin la plus efficace en termes de temps de calcul. Cependant, on peut remarquer la grande différence entre les deux modèles gaussiens. En effet, le modèle à variance fixe permet d'accélérer grandement le calcul, cela venant sûrement de la simplicité de calcul de ses paramètres. En effet, les amplitudes représentent les paramètres les plus faciles à maximiser. De plus, en interdisant les inversions de position des gaussiennes, nous réduisons la singularité du système. Cependant, comme nous le disions précédemment, les échantillons de la base de données peuvent être de contenu très disparate, les graphiques et statistiques complètes sont donc disponibles en annexe.

3.6.3 Intégration dans *Orchidée*

D'après les résultats présentés dans la section précédente, le modèle le plus avantageux est le mélange de gaussiennes à variance fixe. Celui-ci permettant de représenter l'information temporelle, nous pouvons ainsi étendre grandement le système d'orchestration *Orchidée* développé par Carpentier [14]. De cette manière la recherche d'orchestration ne se limiterait plus aux sons purement stationnaires. Cependant

le système ayant été programmé autour de l'hypothèse centrale de sons harmoniques et sans évolution temporelle, nous devons nous heurter à un grand nombre de problèmes. Tout d'abord les descripteurs sont supposés faire la même taille pour tous les sons de la base de données, ce qui a peu de sens pour une modélisation temporelle efficace. De plus, les descripteurs sont tous vectoriels et non matriciels ce qui limite la structuration possible de la modélisation. Un autre problème provient sans doute des fonctions de calcul des combinaisons. Celles-ci se font indépendamment de la cible, ce qui doit être modifié pour permettre d'utiliser les fonctionnalités de modification de durée présentées précédemment. D'autre part, il est impossible d'induire un décalage temporel entre les différents instruments. L'évidence de cette possibilité au vu de l'information temporelle doit cependant être nuancée. En effet l'introduction d'un tel décalage induirait nécessairement une nouvelle combinatoire interne à chaque combinaison, en plus de la reprogrammation de l'algorithme de génération. Nous considérerons donc que les instruments d'une combinaison démarrent de manière synchrone, même si le modèle proposé peut aisément accepter une extension comprenant ce décalage. Enfin, le plus gros problème provient de l'hypothèse harmonique, celle-ci est utilisée pour l'indexation ainsi que le filtrage des sons. A cause de cette hypothèse, toute la structure de l'algorithme n'accepte que des sons ne représentant qu'une seule note. Il est donc pour le moment impossible d'utiliser les glissandos ainsi que les multiphoniques avec les descripteurs existants et le modèle temporel simultanément.

La base de données a donc dans un premier temps été augmentée par des sons à forte variation temporelle. Dans un second temps, l'analyse des modèles présentés dans la section précédente a été réalisée sur l'ensemble des échantillons instrumentaux. Enfin après avoir modifié certaines parties de la structure de l'algorithme, les fonctions permettant d'ajouter le descripteur d'évolution temporelle ont été réalisées pour permettre son utilisation en situation réelle.

3.6.3.1 Expériences

Nous décrivons ici les expériences réalisées avec le système *Orchidée* après l'intégration du modèle temporel. Nous proposons ainsi des expériences de complexité progressive permettant de poser le pied dans le monde du micro-temps jusqu'à observer la nette amélioration dans les orchestrations proposées.

- La première expérience nous permet d'attaquer la temporalité de manière simple en utilisant un son ayant une structure fréquentielle "relativement" fixe mais dont l'amplitude relative varie au cours du temps. Nous utilisons pour ce faire une voix extraite de *Poetry* de Marco Suarez. Cette voix produit une note tenue mais se composant de 3 parties distinctes. D'abord une composante bruitée puis une amplification de l'intensité qui s'achève par un tremolo.
- La deuxième expérience permet de prendre pleine mesure de l'avancée effectuée par l'intégration de la composante micro-temporelle. Nous orchestrans ici des sons à très forte variation temporelle et dont la temporalité pourrait même être considérée comme la caractéristique principale. Nous utilisons pour ce faire un échantillon de la voix de Chewbacca issu du film *La guerre des étoiles*. Ce son présente une modulation caractéristique à la fois sur le plan harmonique et temporel. Il commence ainsi par une partie bruitée très saccadée et se termine par une partie harmonique dont la note se situe environ deux tons sous celle de départ.

3.6.3.2 Comparaison

Nous comparons à présent les résultats à travers les expériences proposées précédemment. Cette comparaison nous permet d'entrevoir le potentiel du descripteur temporel en observant les orchestrations proposées sans la prise en compte de notre modèle puis les orchestrations résultant de l'intégration de nos travaux au sein du système *Orchidée*. Nous avons réalisé un script Matlab permettant de faire l'interface avec le logiciel *LilyPond* et de générer ainsi de manière automatique la partition relative à la solution proposée par le système. Ce script génère pour chaque instrument une portée contenant son nom, son mode de jeu, la note jouée avec son intensité ainsi que les différents signes d'interprétation tels que les crescendo, decrescendo et trilles.

Figure 3.4 displays musical notation for various instruments (Fl, Ob, ClBb, Vn, Va, Vc) across three systems. The notation includes dynamic markings (e.g., *f*, *pp*, *mf*, *ff*, *p*) and articulations (e.g., *ord*, *flatt*, *nonvib*, *legno-tratto*, *pont_ord*, *legno-batt*).

Figure 3.4: Résultats de l'orchestration d'une voix extraite de *Poetry* de Marco Suarez en utilisant uniquement des descripteurs statiques

Figure 3.5 displays musical notation for various instruments (Ob, Vn, Vc, Va, ClBb, Cb) across three systems. The notation includes dynamic markings (e.g., *p*, *mf*, *ff*, *pp*) and articulations (e.g., *vib*, *legno-tratto*, *ord_trem*, *cresc*, *trill-maj2*, *cre_dec*, *art-harm*).

Figure 3.5: Résultats de l'orchestration de la même cible après intégration de notre modèle temporel

The figure displays two side-by-side musical staves for an orchestra, comparing static descriptors (left) with the results after integrating a temporal model (right). The instruments listed on the left are Fl, Ob, ClBb, Bn, Hn, TTbn, BTb, Va, Vc, and Cb. The instruments on the right are Fl, Ob, ClBb, TpC, TTbn, BTb, Vn, Vc, and Cb. The left side shows static descriptors like *ord*, *pp*, *stacc*, *mf*, *ff*, *p*, *mf*, *ord-open*, *f*, *tasto*, *ff*, *pont*, *mf*, and *nonvib*. The right side shows dynamic and expressive descriptors like *cresc*, *pp*, *ff*, *lip-gl-dsc-slow*, *mf*, *cre_dec*, *pp*, *mf*, *pp*, *flatt*, *decresc*, *ff*, *pp*, *flatt*, *ff*, *gl-slw*, *mf*, *trill-min2*, *tr*, *mf*, *trem*, *pp*, and *art-harm*.

Figure 3.6: Comparaison des résultats de l'orchestration de la voix de *Chewbacca* en utilisant uniquement des descripteurs statiques (gauche) et après intégration du modèle temporel (droite)

Comme on peut le voir sur les résultats de la première expérience, lors de l'utilisation de descripteurs statiques uniquement, même si la structure harmonique moyenne est respectée, l'information temporelle est complètement omise. On peut tout de même noter la présence de deux *legno-tratto* qui simulent la partie bruitée mais durant toute l'orchestration résultante. Après intégration du descripteur temporel cependant, nous observons très nettement sur les résultats choisis les différentes parties du son d'origine. En effet la partie bruitée est tour à tour simulée par un violon en *legno-tratto*, un alto en *legno-tratto* puis finalement une clarinette en *harmoniques artificiels*. La modification d'intensité est également présente sous la forme de *crescendo* et *crescendo - decrescendo*. Enfin le tremolo final est également fortement présent dans le premier et troisième résultat. Il est d'ailleurs intéressant de noter que le dernier résultat respecte moins la structure globale de la cible mais se concentre sur l'aspect "vibré" final par la présence d'un tremolo ainsi que d'une trille. Nous confirmons donc ainsi la validité de notre modèle et la nette amélioration que celui-ci apporte au système existant.

La seconde expérience est conçue pour pousser les limites du système et tester le modèle temporel dans des conditions particulièrement difficiles. De par la complexité requise par une telle orchestration, les résultats font appel à plus d'une dizaine d'instruments et une grande variété de modes de jeu. En comparant les deux orchestrations proposées, on s'aperçoit rapidement du potentiel du modèle temporel. En effet, même si le modèle statique imite de manière précise la structure harmonique de la cible, celui-ci ne nous propose qu'une "texture" approximant l'intégralité du son. A l'inverse le modèle temporel en vient presque à "surcharger" l'écriture. Ainsi chaque élément du résultat proposé permet de tenir compte d'une partie spécifique du son d'entrée. Le violon effectuant une *trille*, le violoncelle en *trémolo* ainsi que le trombone et le tuba en *flutterzunge* permettent d'imiter la structure temporelle typiquement saccadée de la cible. Le hautbois et le trombone en *glissando* reproduisent la modification de structure harmonique. Enfin la flûte qui effectue un *crescendo* permet de retrouver la structure harmonique finale.

Les résultats préliminaires de notre modèle sont donc très prometteurs et permettent d'envisager à terme la prise en compte complète de la micro-temporalité dans l'orchestration. D'un point de vue perceptif, la forme temporelle s'est révélée être presque plus importante que la structure harmonique lors des comparaisons subjectives.

Chapter 4

Evolution macro-temporelle

Après la définition d'un modèle micro-temporel nous pouvons envisager une plongée dans l'échelle du temps vers le niveau du macro-temps. Cependant, l'idée d'une continuité entre macro et micro-temps doit sûrement être prise avec la plus grande précaution. Nous devons donc faire face ici à de nombreux problèmes. D'une part, nous ne disposons pas de règles régissant les combinaisons instrumentales et encore moins l'évolution temporelle entre ces différents instruments. D'autre part il n'existe pas de lien direct entre les domaines du micro et macro-temps.

Ainsi notre recherche s'engageant dans la voie du temps, il nous faut abandonner, ne serait-ce que temporairement, l'approche combinatoire, verticale, de l'orchestration pour se concentrer sur une modélisation du temps. Nous choisissons donc dans un premier temps de tenter d'étendre l'approche de l'orchestration par un son cible en permettant à celui-ci d'être un morceau complet ayant différents éléments à orchestrer. En inversant le paradigme, l'orchestration musicale peut être considérée comme un problème de séparation de sources où le but est de trouver la combinaison d'instruments qui explique au mieux la cible. Compte tenu de cette approche, nous utilisons la factorisation en matrices non-négatives pour permettre de décomposer le spectre de la cible en une somme pondérée des instruments choisis. L'objectif est d'abord de découper le spectre dans le temps en unités qui ont une certaine cohérence spectrale. Cette technique permet à la fois de trouver le mélange instrumental 'optimal' pour un élément mais également d'envisager l'utilisation du système dans un cadre de *réorchestration* d'une musique entière. Ainsi grâce à la technique de factorisation de matrices non-négatives nous pouvons calculer rapidement la décomposition sur une base instrumentale. L'ajout des contraintes de parcimonie nous permet de limiter l'effectif instrumental, d'autre part les contraintes de continuité locale nous permettent également d'ajouter une logique temporelle en pénalisant les propositions ayant de trop grandes variations instrumentales (à la fois de notes et de classes d'instruments).

4.1 Factorisation de matrices non-négatives

4.1.1 Philosophie générale

Originellement développée par Paatero et Tapper [50] sous le terme malencontreux de 'factorisation de matrices positives' et rendue populaire par Lee et Seung [42], la factorisation de matrices non-négatives (NMF) est une méthode innovante d'analyse en sous-espace qui vise à trouver une représentation par parties des objets en imposant une contrainte de non-négativité. La factorisation en matrices non-négatives est une méthode générale de décomposition matricielle. La représentation d'un objet (décrit par un vecteur de réels positifs) comme une somme de ces vecteurs de base, correspond alors à l'intuition d'une décomposition par parties. Cette idée est supportée par l'évidence physiologique et psychologique que la perception d'un tout est basé sur une représentation par parties [47]. Le problème imposé par la NMF est donc de trouver une approximation d'une matrice non-négative donnée, $\mathbf{V} \in \mathbb{R}_{m \times n}^+$ comme le

produit de facteurs non-négatifs, $\mathbf{W} \in \mathbb{R}_{m \times k}^+$ et $\mathbf{H} \in \mathbb{R}_{k \times n}^+$

$$\mathbf{V} \approx \mathbf{WH} \quad (4.1)$$

où la matrice $\mathbf{W}$ de taille $m \times k$ contient les vecteurs de base et la matrice $\mathbf{H}$ de taille $k \times n$ contient les poids permettant d'approximer les colonnes correspondantes de la matrice $\mathbf{V}$ comme une combinaison linéaire des vecteurs de base contenus dans $\mathbf{W}$. Habituellement, le nombre de vecteurs de base est choisi de telle manière que $(n + m)k < nm$, permettant ainsi à la matrice de facteur $\mathbf{H}$ de représenter une version compressée de la matrice de données originelle. On peut voir ici que dans notre approche globale se trouvant à la croisée des mondes du signal et de la symbolique que cette technique nous permet d'obtenir une représentation symbolique à partir de données issues du signal.

4.1.1.1 Propriétés fondamentales

Habituellement, le nombre de colonnes de $\mathbf{W}$ et le nombre de lignes de $\mathbf{H}$ sont choisis pour que le produit $\mathbf{WH}$ devienne une approximation de $\mathbf{V}$. La décomposition complète de $\mathbf{V}$ est alors effectuée par le produit des deux matrices $\mathbf{W}$ et $\mathbf{H}$ et l'ajout d'un résidu $\mathbf{R}$, de telle manière que :

$$\mathbf{V} = \mathbf{WH} + \mathbf{R} \quad (4.2)$$

Les éléments de la matrice résidu $\mathbf{R}$ peuvent alors être indifféremment positifs ou négatifs. Quand les matrices $\mathbf{W}$ et $\mathbf{H}$ sont plus petites que $\mathbf{V}$ elles deviennent alors plus faciles à stocker et à manipuler. De plus, la factorisation obtenue n'est pas unique, une matrice et son inverse peuvent être utilisées [?] pour transformer les deux matrices de factorisation par :

$$\mathbf{WH} = \mathbf{WBB}^{-1}\mathbf{H} \quad (4.3)$$

Si les deux nouvelles matrices $\tilde{\mathbf{W}} = \mathbf{WB}$ et $\tilde{\mathbf{H}} = \mathbf{B}^{-1}\mathbf{H}$ sont non-négatives, elles forment alors une autre paramétrisation de la factorisation. La non-négativité de $\tilde{\mathbf{W}}$ et $\tilde{\mathbf{H}}$ s'applique au moins si $\mathbf{B}$ est une matrice non-négative monomiale. Dans ce cas, cela correspondra simplement à une mise à l'échelle et une permutation.

4.1.2 Applications

Cette technique connaît un intérêt croissant dans la communauté musicale car elle permet d'obtenir de bien meilleurs résultats que d'autres techniques de décomposition tels que la SVD [7] et cela en un temps de calcul extrêmement réduit. Ainsi la NMF a été appliquée avec succès à la transcription rythmique [33, 51] et musicale [2, 8, 74, 75], l'analyse multipitch [17, 55, 62], la séparation de source [27, 77, 80, 81], la séparation de la parole [61, 78], le suivi de partition [16], la classification instrumentale [5] et même au débruitage de la parole [83].

La factorisation de matrices est généralement non unique, et un grand nombre de méthodes pour la réaliser ont été développées (comme par exemple l'analyse en composantes principales et la décomposition en valeurs singulières) en incorporant différentes contraintes. La factorisation en matrices non-négatives diffère des autres méthodes car elle impose la contrainte que les matrices $\mathbf{W}$ et $\mathbf{H}$ soient non-négatives, c'est à dire que tous leurs éléments doivent être supérieurs ou égaux à zéro.

4.1.3 Différentes fonctions de coût

Il y a différents types de factorisations en matrices non-négatives. Ces types prennent naissance dans le choix parmi les différentes fonctions de coût pour mesurer la divergence entre $\mathbf{V}$ et le produit $\mathbf{WH}$ et également par la régularisation possible des matrices $\mathbf{W}$ et $\mathbf{H}$. Les paramètres du modèle sont adaptés en minimisant une fonction de coût qui mesure l'erreur entre la factorisation et la matrice donnée en entrée. L'algorithme de factorisation des NMF peut ainsi être écrit sous forme d'un problème d'optimisation

$$\min_{\mathbf{W}, \mathbf{H}} D_{KL}(\mathbf{V} \parallel \mathbf{WH}) \quad (4.4)$$

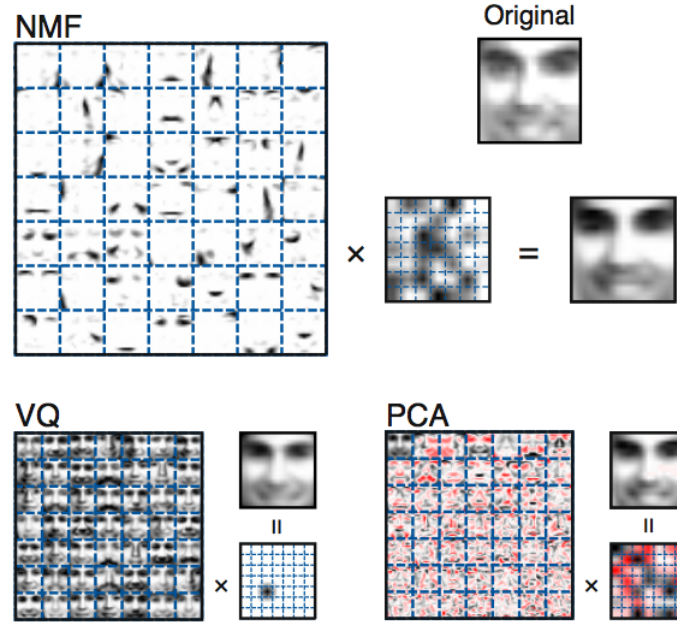


Figure 4.1: Utilisation de l'algorithme NMF (image tirée de [42]) pour la reconnaissance d'images faciales.

Les résultats de l'algorithme sont comparés avec la quantification vectorielle (VQ) et l'analyse en composantes principales (PCA). On peut remarquer que la NMF permet d'obtenir une meilleure représentation intuitive des structures locales.

Deux fonctions de divergence simples étudiées par Lee et Seung [43] sont l'*erreur quadratique* (ou *norme de Frobenius*) et une extension de la *divergence de Kullback-Leibler généralisée* (sa version originelle étant définie pour des distributions de probabilités) qui est la plus régulièrement utilisée :

$$D_{KL}(\mathbf{V} \parallel \mathbf{WH}) = \left\| \mathbf{V} \otimes \log \frac{\mathbf{V}}{\mathbf{WH}} - \mathbf{V} + \mathbf{WH} \right\| \quad (4.5)$$

où $\otimes$ représente un produit élément par élément (aussi connu sous le nom de produit de Hadamard) et la division se fait également par élément.

Chaque mesure de dissimilarité mène à un algorithme différent, tous essayant généralement de minimiser la divergence en utilisant des règles de mise à jour itératives.

Le problème de factorisation en utilisant le critère d'erreur quadratique peut être posé de la manière suivante : étant donné une matrice non-négative $\mathcal{X}$, trouver les matrices $\mathcal{W}$ et $\mathcal{H}$ qui minimisent la fonction :

$$f_{divergence}(\mathbf{W}, \mathbf{H}) = \|\mathbf{V} - \mathbf{WH}\|_F^2 \quad (4.6)$$

4.1.4 Détermination des matrices

Déterminer les matrices $\mathbf{W}$ et $\mathbf{H}$ revient à minimiser la distance (ou fonction de coût) entre la matrice initiale $\mathbf{V}$ et le produit $\mathbf{WH}$ sous les contraintes de non-négativité. Un grand nombre d'algorithmes de mises à jour ont été développés et les équations sont souvent fortement dépendantes du problème. Cependant, ces algorithmes peuvent être divisés en trois classes générales : multiplicatifs, descente de gradient et moindres carrés alternés.

4.1.4.1 Algorithmes multiplicatifs

Dans leur article, Lee et Seung [42] ont originellement proposé de résoudre le problème d'optimisation des NMF en utilisant des mises à jour multiplicatives. Leur algorithme utilise la norme de Frobenius et consiste à initialiser $\mathbf{W}$ et $\mathbf{H}$ aléatoirement, puis à alterner les mises à jour suivantes :

$$\mathbf{H} \leftarrow \mathbf{H} \otimes \frac{(\mathbf{W}^T \mathbf{V})}{(\mathbf{W}^T \mathbf{W} \mathbf{H} + \epsilon_d)} \quad (4.7)$$

$$\mathbf{W} \leftarrow \mathbf{W} \otimes \frac{(\mathbf{V} \mathbf{H}^T)}{(\mathbf{W} \mathbf{H} \mathbf{H}^T + \epsilon_d)} \quad (4.8)$$

Le paramètre ϵ_d est ajouté pour éviter une division par zéro et on choisit généralement $\epsilon_d = 10^{-9}$. Les auteurs utilisent le gradient et la propriété de descente continue pour prouver que leur algorithme converge vers un minimum local, cependant il a par la suite été démontré [25] que cette preuve était fausse. En réalité, la seule conclusion [45] qui puisse être faite sur la convergence de cet algorithme est qu'une convergence s'effectue vers un point stationnaire. Par son statut de premier algorithme connu, l'algorithme multiplicatif est devenu une référence à laquelle les autres algorithmes sont comparés. Cependant, il a été montré à plusieurs reprises que cet algorithme (s'il converge) nécessite un grand nombre d'itérations avant de converger. De plus le coût de calcul par itération est important. Chaque itération nécessite six multiplications de matrices en $\mathcal{O}(n^3)$ et six opérations terme à terme en $\mathcal{O}(n^2)$.

4.1.4.2 Algorithmes de descente de gradient

La seconde classe d'algorithmes est basée sur les méthodes de descente de gradient. Les équations de mises à jour sont alors

$$\mathbf{H} \leftarrow \mathbf{H} - \epsilon_H \frac{\delta f}{\delta \mathbf{H}} = \mathbf{H} - \epsilon_H \mathbf{W}^T (\mathbf{W} \mathbf{H} - \mathbf{V}) \quad (4.9)$$

$$\mathbf{W} \leftarrow \mathbf{W} - \epsilon_W \frac{\delta f}{\delta \mathbf{W}} = \mathbf{W} - \epsilon_W (\mathbf{W} \mathbf{H} - \mathbf{V}) \mathbf{H}^T \quad (4.10)$$

Les paramètres de taille du pas ϵ_H et ϵ_W varient suivant les algorithmes utilisés. Ces algorithmes prennent toujours la direction du gradient négatif, la subtilité provenant du choix des valeurs pour les paramètres de pas. Certains algorithmes [35] initialisent ces valeurs à 1, puis les multiplient par leur moitié à chaque itération. Cependant, il n'y a pas de restriction sur le fait que les éléments des matrices mises à jour deviennent négatifs. Ce problème est généralement contourné par une simple étape de projection par laquelle tous les éléments négatifs sont remis à zéro.

4.1.4.3 Algorithmes des moindres carrés alternés.

La dernière classe d'algorithmes alterne deux étapes de moindres carrés. Cet algorithme a été utilisé dans l'article original de Paatero [50] et exploite le fait que même si le problème n'est pas convexe à la fois pour $\mathbf{W}$ et $\mathbf{H}$, il est convexe soit pour $\mathbf{W}$ soit pour $\mathbf{H}$. Donc si une matrice est donnée, l'autre peut être trouvée par un calcul simple des moindres carrés. L'algorithme consiste donc à trouver de manière alternée les matrices $\mathbf{H}$ et $\mathbf{W}$ qui satisfont les équations matricielles

$$\text{Trouver } \mathbf{H} \text{ tel que } \mathbf{W}^T \mathbf{W} \mathbf{H} = \mathbf{W}^T \mathbf{V} \quad (4.11)$$

$$\text{Trouver } \mathbf{W} \text{ tel que } \mathbf{H} \mathbf{H}^T \mathbf{W}^T = \mathbf{H} \mathbf{V}^T \quad (4.12)$$

Cependant, on peut remarquer que comme l'algorithme de descente de gradient, les matrices mises à jour peuvent contenir des éléments négatifs. Ce problème est également contourné par une étape de projection par laquelle tous les éléments négatifs sont remis à zéro.

4.1.5 Extensions

Il y a différentes manières de trouver les matrices $\mathbf{W}$ et $\mathbf{H}$. Les algorithmes peuvent s'avérer loin d'être idéaux, étant donné qu'ils ne garantissent typiquement que de trouver un minimum local, plutôt qu'un minimum global de la fonction de coût mais dans de nombreuses applications de fouille de données, un minimum local peut être suffisant suivant le but recherché. Cependant pour améliorer les résultats, le problème est généralement étendu pour inclure des contraintes auxiliaires sur $\mathbf{W}$ et $\mathbf{H}$. Cela est fait en général pour compenser des incertitudes sur les données, pour imposer des caractéristiques désirées sur la solution obtenue ou pour imposer un savoir a priori sur les données. La fonction de coût est étendue de manière générale par des termes de pénalité permettant d'imposer les contraintes auxiliaires

$$f_{\text{constraint}}(\mathbf{V}, \mathbf{WH}) = f(\mathbf{V}, \mathbf{WH}) + \alpha C_1(\mathbf{W}) + \beta C_2(\mathbf{H}) \quad (4.13)$$

Ici, $C_1(\mathbf{W})$ et $C_2(\mathbf{H})$ sont les fonctions de pénalité introduites pour imposer certaines contraintes. Les paramètres α et β sont des paramètres de régularisation permettant de définir la balance entre l'erreur d'approximation et les contraintes imposées. Les algorithmes diffèrent principalement par les contraintes incluses dans leur fonction objective et sont présentées ci-après.

4.1.5.1 Contraintes simples

La NMF standard impose les contraintes de non-négativité sur les matrices $\mathbf{W}$ et $\mathbf{H}$, ainsi un vecteur de données peut être approximé par une combinaison additive des vecteurs de base.

$$\min_{\mathbf{W}, \mathbf{H}} D_k(\mathbf{V} \parallel \mathbf{WH}) \quad (4.14)$$

Les contraintes de régularité sont souvent imposées pour lisser les solutions en cas de présence de bruit dans les données. Par exemple le terme

$$C_1(\mathbf{W}) = \|\mathbf{W}\|_F^2 \quad (4.15)$$

pénalise les solutions $\mathbf{W}$ dont la norme de Frobenius est importante. De manière plus générale, il est possible de régulariser les valeurs de $\mathbf{W}$ en imposant la contrainte $C_1(\mathbf{W}) = \|\mathbf{LW}\|_F^2$ avec $\mathbf{L}$ un opérateur de régularisation. Il est également possible d'imposer une contrainte de continuité temporelle sur les colonnes de $\mathbf{H}$ en définissant

$$C_2(\mathbf{H}) = \frac{1}{n} \|(\mathbf{I} - \mathbf{T})\mathbf{H}^T\|_F^2 \quad (4.16)$$

où n représente le nombre total de colonnes dans la matrice de données $\mathbf{V}$ et $\mathbf{T}$ est un opérateur de convolution. Suivant le problème, ces contraintes peuvent améliorer les résultats ou les faire tendre vers une forme particulière souhaitée.

4.1.5.2 Factorisation parcimonieuse

Une des propriétés intéressantes de la version originale des NMF est qu'elles produisent généralement une représentation parcimonieuse des données. Cela peut être utile dans notre application pour éviter les doublons d'octave qui sont généralement inutiles. Cependant cette parcimonie ne peut être contrôlée directement et elle apparaît généralement plus comme un effet de bord que comme un but recherché de l'algorithme. Ainsi, en s'inspirant du codage parcimonieux, Hoyer [35] introduisit une extension parcimonieuse des NMF, permettant de contrôler directement le niveau de parcimonie. Cette extension nous permet d'imposer des résultats utilisant un effectif instrumental minimal. Ce modèle a par ailleurs été utilisé par Cont [18] pour réaliser une détection simultanément multipitch et multi-instruments. Le but des NMF parcimonieuses (SNMF) est d'imposer une contrainte qui permettrait de révéler des caractéristiques locales parcimonieuses sur la matrice de données $\mathbf{V}$. La fonction de coût est ainsi modifiée pour prendre en compte cette parcimonie

$$D_s(\mathbf{V} \parallel \mathbf{WH}) = D_{KL}(\mathbf{V} \parallel \mathbf{WH}) + \lambda \cdot \text{sparse}(h_i) \quad (4.17)$$

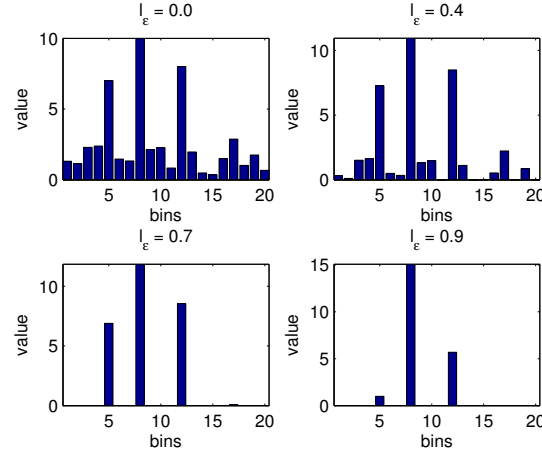


Figure 4.2: Un vecteur synthétique (en haut à gauche) et les résultats d'une projection parcimonieuse pour des valeurs $l_\epsilon = 0.4, 0.7, 0.9$

où λ est une constante positive définissant le partage entre la quantité de parcimonie requise et l'exactitude de la reconstruction. Cette fonction de coût reflète combien d'énergie de la solution est contenue dans un minimum de composantes et donc pénalise les solutions avec de nombreux facteurs différents de zéro. Dans son article, Hoyer [35] propose d'utiliser une mesure de parcimonie basée sur la relation entre les normes l_1 et l_2

$$sparse(h_i) = \frac{\sqrt{I} - (\sum |h_i|) / \sqrt{\sum h_i^2}}{\sqrt{I} - 1} \quad (4.18)$$

où I représente la dimensionnalité de $\mathbf{H}$. Nous utilisons la mesure de parcimonie proposée par Cont dans [18] et qui est définie par

$$sparse(h_i) = \frac{\sqrt{I} - \sum \tanh(|h_i|^2) / \sqrt{\sum h_i^2}}{\sqrt{I} - 1} \quad (4.19)$$

La procédure de projection peut être exprimée de la manière suivante. On commence par projeter le vecteur donné h_i sur l'hyperplan l_ϵ

$$s_i = h_i + \frac{(l_\epsilon - \sum \tanh(h_i^2))}{I} \quad (4.20)$$

Ensuite, on projette sur le point le plus proche de l'hypersphère jointe. Il faut donc résoudre l'intersection de la projection avec l'hyperplan l_2 . Cela est fait en résolvant une équation quadratique pour trouver le facteur α tel que $s = m + \alpha(s - m)$ satisfasse la contrainte l_2 . Comme on peut le voir dans l'équation précédente, certains éléments du vecteur projeté peuvent être négatifs. Si tel est le cas, ces éléments sont ramenés à zéro.

4.1.5.3 Factorisation locale

Pour imposer des contraintes concernant la localisation spatiale et conséquemment révélant des caractéristiques locales dans la matrice de données $\mathbf{V}$, la NMF locale incorpore 3 contraintes additionnelles au problème standard :

1. Minimiser le nombre de composantes de base représentant $\mathbf{V}$.
2. Les différents vecteurs doivent être les plus orthogonaux possible.

3. Retenir les composantes donnant l'information la plus importante.

Les contraintes ci-dessus sont exprimées en modifiant la fonction de divergence de la manière suivante

$$D_m(\mathbf{V} \parallel \mathbf{WH}) = D_k + \alpha \sum_{i=1}^r \sum_{j=1}^r u_{ij} - \beta \sum_{i=1}^r \sum_{j=1}^r q_{ij} \quad (4.21)$$

α, β étant des constantes, $[u_{ij}] = \mathbf{W}^T \mathbf{W}$ et $[q_{ij}] = \mathbf{H}^T \mathbf{H}$

4.1.5.4 Factorisation discriminante

On conserve ici les contraintes de l'algorithme original en tentant d'améliorer la localisation des vecteurs de base de la NMF locale et on essaye d'améliorer la précision en incorporant des informations de discrimination de classes. Deux contraintes supplémentaires sont introduites :

1. Minimiser la matrice de dispersion intra-classe $\mathcal{S}_w$.
2. Maximiser la matrice de dispersion inter-classes $\mathcal{S}_b$.

Les contraintes ci-dessus sont exprimées en modifiant la fonction de divergence de la manière suivante

$$D_m(\mathbf{V} \parallel \mathbf{WH}) = D_k + \alpha \sum_{i=1}^r \sum_{j=1}^r u_{ij} - \beta \sum_{i=1}^r \sum_{j=1}^r q_{ij} + \gamma \mathcal{S}_w - \delta \mathcal{S}_b \quad (4.22)$$

avec γ, δ étant des constantes

4.1.5.5 Factorisation pondérée

Enfin d'autres modèles peuvent être pris en compte comme la weighted NMF [30] qui permet d'accorder une importance plus ou moins importante aux différentes données. Pour inclure la pondération [31] dans le problème d'approximation des NMF, celui-ci peut être étendu de la manière suivante

$$\mathbf{V} \approx \sum_{i=1}^k \sigma_i w_i h_i^T \quad (4.23)$$

où les éléments non-négatifs σ_i sont les facteurs de poids de la combinaison linéaire. Il est ainsi possible d'inclure une matrice de pondération dans le problème d'approximation et les différents algorithmes de mises à jour des facteurs peuvent être adaptés pour inclure cette pondération. Les fonctions de coût sont tout d'abord modifiées pour prendre en compte la pondération. On obtient ainsi la distance euclidienne pondérée

$$F_P(\mathbf{V}, \mathbf{WH}) = \frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|_P^2 = \frac{1}{2} \sum_{ij} [\mathbf{P} \otimes (\mathbf{V} - \mathbf{WH}) \otimes (\mathbf{V} - \mathbf{WH})]_{ij} \quad (4.24)$$

et la divergence de Kullback-Leibler pondérée

$$D_{KL_P}(\mathbf{V} \parallel \mathbf{WH}) = \sum_{ij} \left[\mathbf{P} \otimes \left(\mathbf{V} \otimes \log \frac{\mathbf{V}}{\mathbf{WH}} - \mathbf{V} + \mathbf{WH} \right) \right]_{ij} \quad (4.25)$$

La distance euclidienne pondérée $\frac{1}{2} \|\mathbf{V} - \mathbf{WH}\|_P^2$ est non-augmentante par les règles de mises à jour

$$\mathbf{H} \leftarrow \mathbf{H} \otimes \frac{[\mathbf{W}^T (\mathbf{P} \otimes \mathbf{V})]}{[\mathbf{W}^T (\mathbf{P} \otimes (\mathbf{WH}))]} \quad (4.26)$$

$$\mathbf{W} \leftarrow \mathbf{W} \otimes \frac{[(\mathbf{P} \otimes \mathbf{V}) \mathbf{H}^T]}{[(\mathbf{P} \otimes (\mathbf{WH})) \mathbf{H}^T]} \quad (4.27)$$

La distance est invariante par ces mises à jour si $\mathbf{W}$ et $\mathbf{H}$ sont à un point fixe de la distance.

4.2 Notre approche

Nous commençons par définir des unités (ou tranches) temporelles qui présentent une *cohérence* spectrale. Cette approche permet à l'entrée cible d'être une chanson ayant plusieurs éléments à orchestrer séparément. De cette façon, notre système peut être utilisé dans une optique de réorchestration, où le compositeur cherche à obtenir le même effet timbral mais par exemple avec un effectif instrumental réduit. Comme nous voulons obtenir des unités ayant une cohérence spectrale, nous utilisons le *flux spectral* comme critère pour délimiter les unités de temps. Nous commençons donc par calculer le flux spectral qui mesure la variation de l'amplitude dans chaque bande de fréquence et est défini par

$$FS(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} H(|X(n, k)| - |X(n-1, k)|) \quad (4.28)$$

où $H(x) = \frac{x+|x|}{2}$ est la fonction de redressement en demi-onde. Cette fonction de détection montre le contraste entre les fenêtres successives de la STFT, ce qui est notre objectif principal puisque nous cherchons des unités spectralement consistantes. Il nous suffit ensuite d'effectuer une sélection de pic sur la fonction pour obtenir la position des *onsets* dans le signal d'entrée.

Notre but est alors de trouver le mélange instrumental qui approxime au mieux chaque élément la cible. On peut clairement voir ici l'analogie entre notre problème et l'algorithme NMF en réécrivant notre stratégie générale de la manière suivante

$$\mathbf{V} = \sum_{i=1}^I \mathbf{H}_i \mathbf{W}_i + \mathbf{R} \quad (4.29)$$

où I est le nombre d'instruments disponibles et $\mathbf{V}$ est la matrice d'entrée représentant la cible. Chaque colonne représente une unité temporelle et contient le résultat de l'analyse par une transformée à Q constant. Dans notre système, la matrice de vecteurs de base $\mathbf{W}$ est fixe et contient les modèles instrumentaux. Enfin, la matrice $\mathbf{R}$ est le résidu qui permet d'obtenir une reconstruction parfaite. Comme la matrice $\mathbf{W}$ est fixée, il nous faut uniquement trouver la matrice de facteur $\mathbf{H}$ correspondante pour reconstruire au mieux la cible donnée en entrée. Nous utilisons dans notre application l'algorithme de descente de gradient dont les règles de mises à jour sont

$$\mathbf{H} \leftarrow \mathbf{H} - \epsilon_H \mathbf{W}^T (\mathbf{W}\mathbf{H} - \mathbf{V}) \quad (4.30)$$

où le paramètre ϵ_H représente la taille du pas de recherche.

Nous utilisons la projection parcimonieuse proposée par Cont [18] telle que nous l'avons présentée au chapitre précédent. L'ajout des contraintes de parcimonie nous permet ainsi de limiter l'effectif instrumental. Cependant notre contribution principale consiste en une approche "alternée" de la parcimonie. Ainsi nous effectuons à chaque itération une première descente de gradient permettant de diminuer l'erreur de reconstruction sans tenir compte de la parcimonie puis nous effectuons une seconde descente de gradient, cette fois dépendante de la parcimonie pour obtenir un résultat final satisfaisant les contraintes de parcimonie requises. Cette approche nous permet à la fois d'accélérer grandement le calcul mais aussi d'augmenter la convergence en évitant le blocage trop rapide dans un minimum local. Cette amélioration s'explique par le fait que le résultat induit par la projection parcimonieuse est grandement dépendant de l'initialisation. Pour comparer notre approche à celle qui n'utilise pas cette alternance, nous doublons le nombre d'itérations de l'algorithme original pour effectuer une comparaison équitable. L'évolution de la fonction de divergence est présentée dans la figure suivante

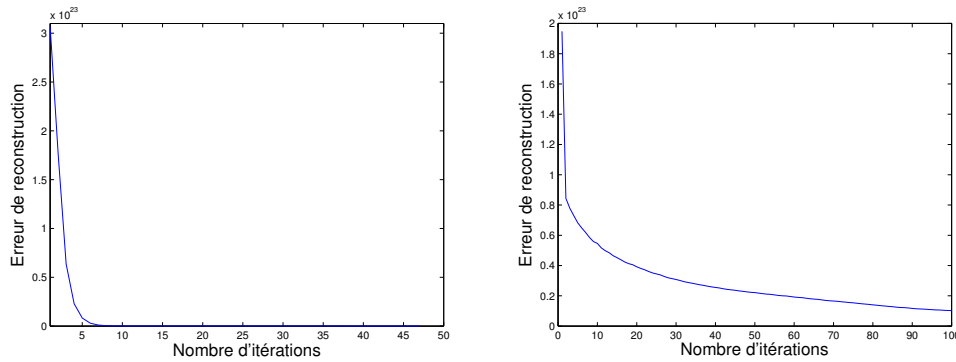


Figure 4.3: Comparaison entre l'approche alternée (gauche) et l'algorithme original (droite) en termes d'évolution de la fonction de divergence et donc de convergence

Nous observons bien sur cette figure que l'approche alternée permet d'accélérer grandement la vitesse de convergence de l'algorithme à la fois en termes de nombre d'itérations mais également en temps de calcul.

4.3 Résultats

4.3.1 Validation

Une première étape de validation consiste à vérifier la capacité de l'algorithme à retrouver les éléments d'une mixture réalisée à partir d'éléments de la base de données. Nous obtenons ainsi une première mesure objective de vérification de l'algorithme. La validation se déroule de la manière suivante. Nous commençons par sélectionner un sous-ensemble de la base de données, permettant ainsi de comparer les performances de l'algorithme sur des tailles variables, notamment en fonction du nombre d'instruments présents. Nous effectuons ensuite une sélection aléatoire de sons présents dans la base pour construire une mixture synthétique aléatoire dont le nombre d'instruments varie également. L'algorithme de factorisation est ensuite appliqué avec la mixture comme donnée d'entrée et le sous-ensemble de la base de données comme matrice de vecteurs de base. Cette procédure est répétée 20 fois pour chaque ensemble de paramètres pour permettre d'obtenir une validation non biaisée par l'initialisation.

Pour mesurer les résultats de l'algorithme de factorisation, nous utilisons les mesures de statistiques bien connues que sont la *précision*, le *recall* et la *F-measure*. Cependant même si ces mesures nous permettent d'évaluer la qualité de l'algorithme, celles-ci sont à nuancer par le fait qu'elles ne prennent pas en compte la valeur des facteurs.

Si l'on analyse les résultats de l'algorithme (disponibles en annexes), on voit que le recall reste relativement constant en fonction du nombre de sons présents dans la mixture, cependant les performances baissent de manière significative en fonction de celui-ci. L'algorithme est donc moins performant pour retrouver une mixture complexe et ceci presque indépendamment de la taille de la base de données. Quand on s'intéresse à la précision, on remarque que celle-ci reste relativement stable en fonction de la taille de la base de données. Ce résultat s'explique facilement étant donné qu'une base de données plus grande implique donc une plus grande possibilité de choix et donc un plus grand nombre de faux positifs. On remarque d'ailleurs que les performances baissent grandement suivant la taille de la base de données. Lorsqu'on compare ces résultats avec ceux d'un facteur de parcimonie inférieur (disponibles en annexe), on remarque que les performances en termes de précision s'effondrent, ceci étant dû à la plus grande "permissivité" de l'algorithme en termes de choix. Cependant les performances en termes de recall sont légèrement supérieures pour un facteur de parcimonie inférieur. Les figures de statistiques et temps de calcul sont disponibles en annexes.

4.3.2 Critiques

Malgré les résultats probants de cette première approche du macro-temps, il convient d'émettre un certain nombre de critiques qui sont apparues au fil des développements entrepris.

Tout d'abord même si le principe de décomposition présente de nombreux attraits en termes de temps de calcul, par l'utilisation d'un seul critère pour la comparaison de timbres, l'algorithme ne peut que proposer une solution unique qu'il juge optimale. Ainsi malgré une réitération de l'algorithme, celui-ci proposera presque toujours la même orchestration (des variations pouvant apparaître en fonction de l'initialisation), ce qui limite l'intérêt artistique d'une telle approche. De plus, cet algorithme orchestre chaque élément sans prendre compte de l'ensemble. Il n'apparaît donc pas de réelle logique temporelle d'évolution globale, ce qui semble indispensable pour la construction d'un discours musical. Enfin, mis à part le coefficient de parcimonie qui permet de limiter plus ou moins l'effectif instrumental, il n'existe pas de réel contrôle sur le déroulement de l'algorithme et donc a fortiori sur les solutions générées.

Les résultats satisfaisants obtenus lors de la phase de validation permettent d'envisager l'utilisation d'une telle technique pour la transcription musicale ou l'analyse de scènes auditives. Cependant, toutes les raisons que nous venons d'énoncer, nous amènent à conclure que la factorisation de matrices non négatives n'est pas encore actuellement suffisante pour restituer toute la subtilité de la question de l'orchestration.

Chapter 5

Conclusion

Nous avons abordé l’orchestration sous un jour nouveau en étudiant la composante temporelle suivant ses deux facettes de micro et macro-temporalité. Grâce à un modèle instrumental, nous pouvons prendre en compte le timbre comme une structure spectro-temporelle. Celui-ci permet à la fois de modéliser de manière précise les propriétés spectrales des instruments mais également d’inférer la structure d’un son de durée modifiée. En intégrant nos résultats au système d’orchestration *Orchidée*, nous avons permis d’ajouter la temporalité à l’approche existante. Celle-ci permet d’ouvrir la voie à un ensemble de sons jusqu’alors inaccessibles mais surtout de générer des orchestrations valides et utilisables dans une situation de concert. En comparant les orchestrations générées avec et sans notre modèle temporel nous avons pu observer la très nette amélioration du système existant.

Nous avons tenté dans un second temps d’utiliser une première approche pour aborder le problème du macro-temps. De par la complexité relative d’un tel problème nous avons temporairement laissé le problème combinatoire en retrait. Cette approche nous a permis de générer des résultats préliminaires et d’imaginer l’utilisation d’un tel système dans un cadre de réorchestration. Cependant cette approche manque encore à ce jour de contrôle et de logique. L’orchestration reste donc un sujet ouvrant de nombreuses questions qui sont toujours à étudier.

5.1 De la complexité de l’orchestration

La complexité de l’orchestration se décline selon trois axes : complexité de la perception du timbre, complexité combinatoire des mélanges instrumentaux, complexité des structures temporelles. La liste des problèmes ouverts par l’orchestration est quasiment infinie, nous tenterons cependant de mettre en exergue certains points suivant les différents axes de complexité.

5.1.1 Complexité combinatoire

Après réflexion, on s’aperçoit que le processus relatif à l’orchestration, s’il a un objectif précis, n’a pas de point de départ systématique. Et c’est justement ce point de départ, cette cible qui fixe le niveau de complexité combinatoire auquel nous aurons à faire face. Dans un premier temps, à la question déjà très réductrice de savoir comment produire un timbre avec un orchestre donné, s’est substituée celle encore plus restreinte de chercher, dans un ensemble d’échantillons de sons instrumentaux, une ou plusieurs combinaisons d’éléments se rapprochant au mieux d’un son cible fourni par l’utilisateur. Cette notion de cible sonore n’est que la simplification la plus élémentaire d’une question qui reste à étudier : peut-on réifier l’intention orchestrale d’un compositeur ? L’étude de la question séculaire de l’intention du compositeur dans la création musicale, par une concertation avec de nombreux artistes permettrait ainsi de mettre en place de nouvelles interfaces de contrôles et d’expressivité musicale. Toute une gamme de techniques d’interaction musicale pourrait découler de la question de la composition à la fois horizontale et verticale. La formulation générale du problème elle-même, par l’utilisation d’un son cible, peut être

considérée comme une grande simplification (par une sur-spécification) des questions soulevées par l’orchestration. Il ne s’agit là que d’une expérience en laboratoire étant donné que le compositeur est rarement en possession d’un tel son. Le but de l’orchestration est généralement de *produire* et non *re-produire* des mélanges instrumentaux. Ainsi, une direction intéressante pour les travaux futurs serait de contourner cette approche “bas niveau” par la définition d’une cible abstraite pour le système.

Abstraire la cible

Le problème principal est de trouver un formalisme adéquat pour la définition d’une cible abstraite. Après plusieurs entretiens avec des compositeurs, une approche intéressante est apparue dans la notion d’esquisse. Les compositeurs comptent généralement sur un premier croquis d’une pièce musicale. Ce schéma définit l’intention principale du compositeur et devrait offrir un cadre large d’investigation pour l’orchestration. Cependant il faut d’abord analyser des esquisses pour savoir à la fois si celles-ci ne relèvent pas d’un processus trop subjectif et en même temps comprendre à quel point le compositeur “pense l’orchestre”. Pour ce problème de génération d’une cible sonore, une approche différente serait l’utilisation d’un ensemble de caractéristiques correspondant au mieux aux attentes du compositeur. Il serait possible d’essayer de construire un système de spécification de cible abstraite, utilisant tous les outils de la CAO pour essayer de converger vers une idée abstraite de timbre mais qui ne se réduit pas à un exemple sonore. Toutefois même si la cible sonore peut être vue comme une sur-spécification, une approche par descripteurs serait en contrepartie une sous-spécification du problème de l’orchestration.

La combinatoire du temps

Comme nous le disions précédemment, le modèle proposé peut aisément accepter une extension comprenant un décalage temporel entre les différents instruments. L’évidence de cette possibilité au vu de l’information temporelle induit nécessairement une nouvelle combinatoire interne à chaque combinaison. Etant donné que le système est basé sur l’exploration génétique, il devrait être possible d’ajouter au génome de chaque combinaison une information sur le point de départ de chaque instrument dans la combinaison. Pour optimiser cette information, il serait possible d’utiliser un algorithme de reconnaissance de formes qui nous permettrait de savoir à quelle position l’instrument minimiserait sa divergence avec la cible.

5.1.2 Complexité timbrale

Comme nous l’avons longuement explicité dans la section 2, la complexité du timbre se décline sous différents axes. Cependant nous pouvons noter ici certaines des lacunes dans notre connaissance actuelle du timbre et de ses multiples facettes.

Modèle instrumental

Une modélisation plus grande du timbre semble indispensable mais l’augmentation du nombre de dimensions implique forcément un risque de “sur-apprentissage”. Il faudrait alors également se tourner vers une inférence automatique des critères grâce à une description plus complète du timbre. Il deviendrait donc inutile d’accumuler les critères de décision si ceux-ci ne permettent pas de distinguer perceptivement les solutions.

De nombreuses perspectives sont ouvertes par le modèle proposé. Dans un premier temps, celui-ci pourrait être optimisé en utilisant en algorithme de peak tracking puis en modélisant l’enveloppe de chaque partiel en considérant la fréquence constante. Ceci nous permettrait une dissociation plus fine de l’évolution temporelle en amplitude et en fréquence. De plus, il apparaît évident au vu des méthodes de modification de durée proposées que d’autres inférences pourraient être faites sur les structures temporelles des différents instruments. En effet un des désavantages majeurs du système existant est de limiter son savoir instrumental aux échantillons présents dans la base de données. Un violon ne peut-il donc effectuer des crescendos que sur quelques notes de sa tessiture ? Il est évident qu’à partir des informations existantes on pourrait inférer la structure des modes tels que le *crescendo*, *decrescendo*,

tremolo et *glissando* pour une variété presque infinie de notes et de nuances. Cependant, de telles inférences ne seraient pertinentes que si elles suivent des contraintes réalistes d'instrumentation. Ainsi l'amélioration de la connaissance sur le savoir instrumental permettrait d'imaginer la mise en place d'un outil d'instrumentation particulièrement recherché par les compositeurs.

On peut noter également l'importance de l'attaque qui semble parfois plus discriminante que la partie entretenue. Ainsi, si on ne s'intéressait qu'à l'attaque, il deviendrait possible de définir des typologies d'attaques représentant différentes formes d'attaques. Les modes d'attaques sont sûrement plus puissants et recherchés aujourd'hui que les avancées sur la partie harmonique.

Phénomènes d'émergence

Les phénomènes d'émergences (l'apparition dans le spectre d'éléments qui n'appartiennent à aucun des instruments d'origine) sont une des composantes de complexité les plus inexplorées du timbre. Certains d'entre eux comme la phase, le masquage ou encore la rugosité ont déjà été explicités et en grande partie inclus dans le savoir scientifique. Ils engendrent alors d'autres questions comme celle de l'efficacité spectrale ou de la balance instrumentale, permettant aux instruments de ne pas être masqués.

On est alors en droit de se demander si des phénomènes plus complexes peuvent enfin être approchés de manière scientifique. Le phénomène de fusion instrumentale est par exemple régulièrement réduit à une correspondance d'harmoniques. Pourtant le problème de fusion est aussi lié à la culture et même s'il s'agit en partie d'un problème de correspondance, les phénomènes de masquage sont également à prendre en compte. Du côté de la psychoacoustique et de la perception du timbre, les phénomènes de fusion entre attaque et résonances sont également à investiguer. En orchestration, les sons non entretenus sont fréquemment employés pour leurs attaques rapides. Dès le début de XX^e siècle, les compositeurs ont créé des timbres nouveaux en les associant aux vents ou aux cordes jouées à l'archet. Le caractère inouï de ces sonorités résulte de leur aspect à la fois percussif et soutenu : ne s'apparentant à aucun instrument connu, elles sont très déroutantes pour l'écoute qui les perçoit comme un timbre "unifié". L'intégration de telles techniques passe donc par une compréhension des phénomènes de couplage entre une attaque rapide produite par une catégorie d'instruments et une résonance assurée par une autre. Une modélisation des unissons est également nécessaire mais elle ne se réduit pas pour autant à l'enregistrement et à l'analyse de ces sons. Enfin l'apprentissage du fonctionnement de tels phénomènes permettrait la définition d'un système de raisonnement qualitatif composé d'un ensemble de termes pertinents pour l'orchestration (tessiture, volume, texture, énergie, fusion, diffraction, profondeur). Il est par ailleurs intéressant de noter qu'une grande partie du vocabulaire inhérent à l'orchestration provient directement de ces phénomènes.

5.1.3 Complexité temporelle

Les aspects temporels du timbre apportent une nouvelle dimension de complexité, où se conjuguent verticalité et horizontalité de l'écriture. L'écriture verticale désigne l'agencement des différentes voix à un instant donné, là où l'écriture horizontale concerne le développement dans le temps d'une seule voix. Les deux échelles d'études que nous avons dégagées permettent d'entrevoir les différents niveaux de complexité. Le niveau micro-temporel concerne les modèles instrumentaux dont nous avons déjà parlé dans la section précédente. Le niveau macro-temporel concerne l'évolution plus globale de l'orchestration ou comme le disait Dufourt de la *gestion de l'inertie*. Ce niveau est une condition nécessaire pour envisager l'orchestration de séquences polyphoniques.

Vers le macro-temps

Ainsi deux modèles de macro-temps peuvent se dégager. Un modèle temporel pour les cibles "segmentées" qui consisterait à découper les cibles en plusieurs passages durant lesquels le timbre reste stable pour trouver des solutions pour chaque partie puis étudier les articulations entre les différentes parties. L'approche prise par notre système et tous ceux existant est typiquement berliozienne, l'addition de sons indépendante de l'écriture. Une autre approche intéressante serait de définir des chemins entre des sons, trouver ainsi le lien qui peut unir deux orchestrations verticales en un flux horizontal. De

l'autre côté, un modèle temporel (très long terme) pour les cibles "articulées" serait constitué de cibles dont le timbre évolue de manière continue. Il est impossible de considérer l'orchestration dynamique comme une simple extension du modèle statique, où l'évolution au cours du temps ne serait rien d'autre qu'une succession de "segments". Il est impératif au contraire d'obtenir une modélisation temporelle des *articulations* instrumentales.

La similarité temporelle

Comme nous le disions, il semble important de définir une méthode de reconnaissance d'un son par aspect temporel. Cet indice pourrait tirer profit des modèles gestaltistes ou de reconnaissance de formes issus des sciences cognitives. Un tel indice permettrait de comparer une *forme globale temporelle* plutôt qu'une divergence point-par-point telle que celle que nous utilisons actuellement. Les méthodes temporelles et cognitivistes dans l'élaboration d'un indice de similitude sont certainement moins "propres" scientifiquement que les méthodes spectrales mais ce sont celles qui pourront présenter le plus d'innovations.

5.2 Axes émergents

L'objectif à long terme étant de créer un système informatique complet d'aide à l'orchestration, il est nécessaire de dégager les axes émergents de notre problématique. En complément de l'intérêt artistique, l'orchestration se présente comme un cas d'étude idéal reliant diverses disciplines scientifiques. Le travail accompli jusqu'à maintenant ne représente qu'une infime partie des questions soulevées par l'orchestration. Ces questions recouvrent, entre autres, les domaines actuels de l'acoustique, du traitement du signal, de la combinatoire, ainsi que de la psychoacoustique. Chacune des disciplines concernées pourra ainsi tirer profit des questions sous-jacentes à l'orchestration. Par exemple, d'un point de vue psychoacoustique, l'étude des notions propres à l'orchestration, telles que la fusion, le volume, la profondeur, entre autres, reste inexistante de nos jours. Au niveau informatique, les diverses techniques appliquées aux domaines symbolique et du traitement du signal devront coexister, et éventuellement donner naissance à des techniques hybrides inespérées. Par exemple, le contrôle des systèmes issus de la recherche opérationnelle par des langages de programmation nous fait croire à la création des nouveaux paradigmes expressifs combinant l'extraction sémantique automatique avec le contrôle pensé de la programmation. Bien que ces divers domaines restent encore des terrains à conquérir, il est possible de passer à un niveau supérieur de modélisation, à savoir, la mise en rapport des ces connaissances pour mieux comprendre une notion qui malgré son rôle central dans de nombreux traités et études, reste un savoir empirique.

Du signal à la symbolique

L'orchestration permet d'aborder pleinement une discipline de l'écriture qui jusqu'alors s'était toujours soustraite à la formalisation, et laisse entrevoir la possibilité d'une compréhension plus fine des rapports entre deux mondes hétérogènes, celui de la partition comme système symbolique et celui du timbre comme réalité sonore. Une piste intéressante serait la reconnaissance d'orchestration qui consisterait à prendre un modèle d'orchestration et décrire celui-ci par l'analyse pour pouvoir ainsi proposer une orchestration similaire. Une approche pertinente pour remédier à ce problème consisterait à utiliser des algorithmes d'apprentissage automatique sur des orchestrations existantes. Tout d'abord une double base de données des orchestrations représentées par des informations symboliques (partitions et instruments correspondants) aussi bien que par des caractéristiques spectrales serait mise en place. Ensuite, le système pourrait analyser et apprendre les relations entre les deux ensembles d'informations. Il est vraiment intéressant d'imaginer qu'une telle approche pourrait combler le fossé entre les mondes de la symbolique et du signal qui sont, jusqu'à présent, profondément séparés.

Il ne faut donc certainement pas oublier l'orchestration classique. C'est ainsi que l'apprentissage à base d'exemples permettrait de trouver des explications sur les combinaisons instrumentales en passant par un repos intellectuel sur des bases solides. Citons "*Les Noces*" de Stravinsky (oeuvre écrite après

le sacre), pour laquelle il lui a fallu 10 ans pour trouver l'orchestration adéquate. Le travail donne donc les moyens de l'idée. Pour trouver quels sont les aspects techniques fondamentaux de l'orchestration on peut passer, par exemple, par l'étude détaillée de chaque composante de l'orchestre. Un autre travail sera celui d'extraire, de tout le corpus d'oeuvres existantes, certains traits typiques d'orchestration, servant un dessein précis. On touche ici à une étude menant à l'identification, pour chaque compositeur, d'une approche personnelle de l'art d'orchestrer. Enfin, couplant les deux axes précédents, on pourrait extraire des généralités relevant de l'ensemble et considérer l'orchestre et le rendu de celui-ci comme émergeant de ses composantes.

Ecrire l'électronique

D'autres pistes sont à explorer, en particulier ce que nous appelons l'orchestration mixte. Celle-ci consiste en l'augmentation par des sons synthétisés d'une orchestration faisant appel à des instruments traditionnels. En généralisant la question de l'orchestration en celle des mélanges sonores quelle que soit l'origine des sons utilisés, ce cas d'étude pourrait servir d'environnement de création sonore pour des parties électroniques dans le cadre de musiques mixtes. Les contraintes d'effectif n'ont alors plus cours pour l'écriture des parties électroniques : les sons sont mélangeables à volonté, et dans toutes les proportions de volume possibles. Autrement dit, la représentation même des orchestrations est dans ce cas à repenser.

Du micro au macro-temps

Par dessus tout, l'orchestration soulève la question de l'existence d'une continuité entre micro et macro-temps. Un raffinement de l'analyse temporelle devrait prendre en compte la *diversité de la nature* des différentes couches ou échelles de temps. Ce qui est sous-entendu ici est un état d'interaction entre ces temporalités de niveaux différents d'organisation. Dans tous les cas, une structuration du micro-temps du timbre n'est pas sans conséquence pour le macro-temps et inversement. Une corrélation non-linéaire des dimensions sous-tend cette problématique. Il existe cependant des cas de corrélation où une stratégie compositionnelle pourrait diriger la cohérence formelle du discours musical.

Le processus compositionnel

L'orchestration est donc aussi une question de formes. L'idée de formes d'orchestration prend tout son sens si elle est vue par un ensemble de données symboliques ou courbes de rugosité, de fusion. D'autre part, l'orchestration est presque toujours liée à la composition dans un processus indivisible. En effet, habituellement les compositeurs écrivent directement pour l'orchestre, donc la symbolique de la composition devrait être connectée au processus d'orchestration. Par l'intégration de la composante macro-temporelle d'une forme désirée, la relation entre l'écriture et le timbre pourrait être précisée dans le cadre d'une cible orchestrale.

Ces questions telles qu'elles sont posées ici, ne concernent pas uniquement le champ du traitement du signal mais révèlent une épistémologie de la composition musicale. Les réponses possibles sont reliées aux concepts conservés comme pertinents, agissant comme un substrat d'une pensée musicale donnée. A chaque niveau du processus compositionnel, nous devons nous occuper de la nature des outils employés et décider par quelles interactions nous voulons composer. Car finalement, s'il est quelque chose à retirer de ce travail, c'est qu'un système d'orchestration ne peut amener autre chose qu'une stimulation d'un génie nécessaire, par le biais, non pas du plagiat, mais de l'apport du savoir et de la pratique existants.

Chapter 6

Annexes

Il est sans importance pour chercher la vérité de savoir d'où provient une idée.
- Bouddha, Majjhima Nikaya.

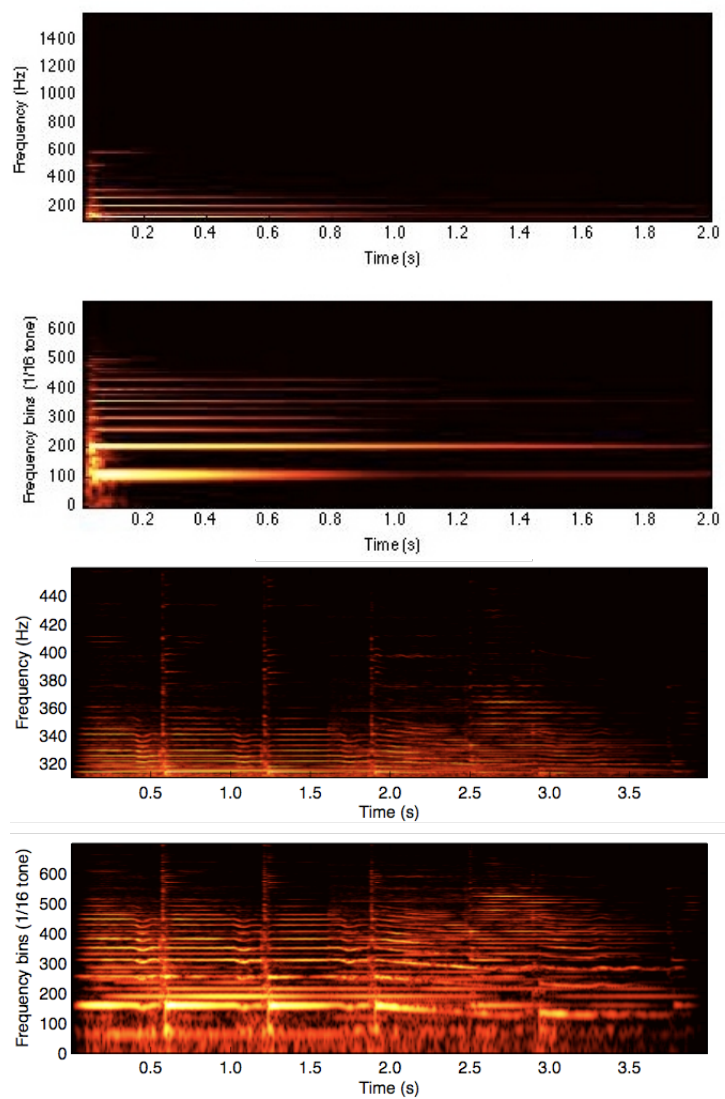


Figure 6.1: Comparaison entre transformée de Fourier et transformée à Q constant pour une note de piano (haut) et une séquence de sitar (bas)

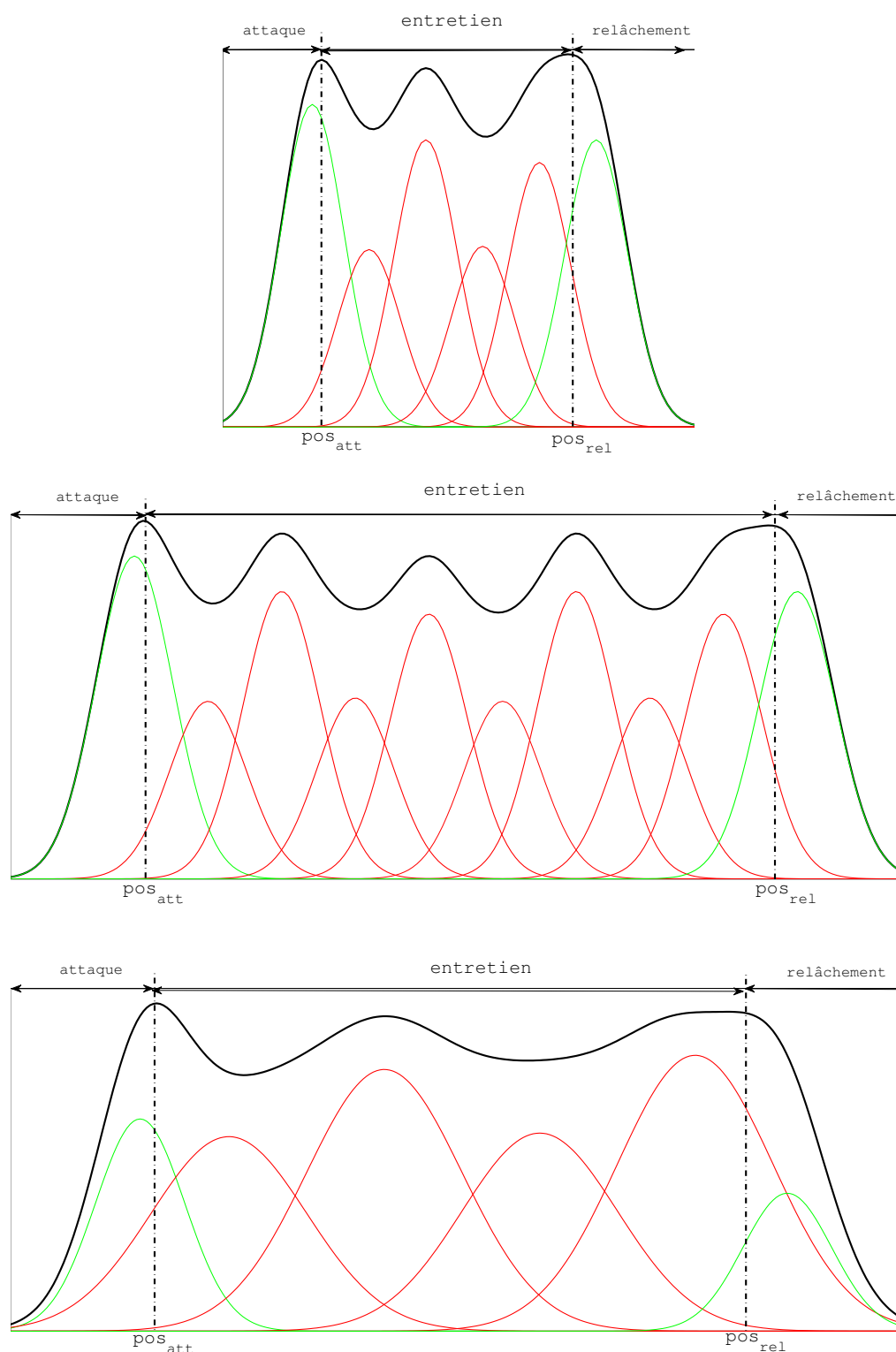


Figure 6.2: Effet d'une modification de durée pertinente par répétition et comparaison avec la dilatation moins pertinente vis-à-vis des données du modèle

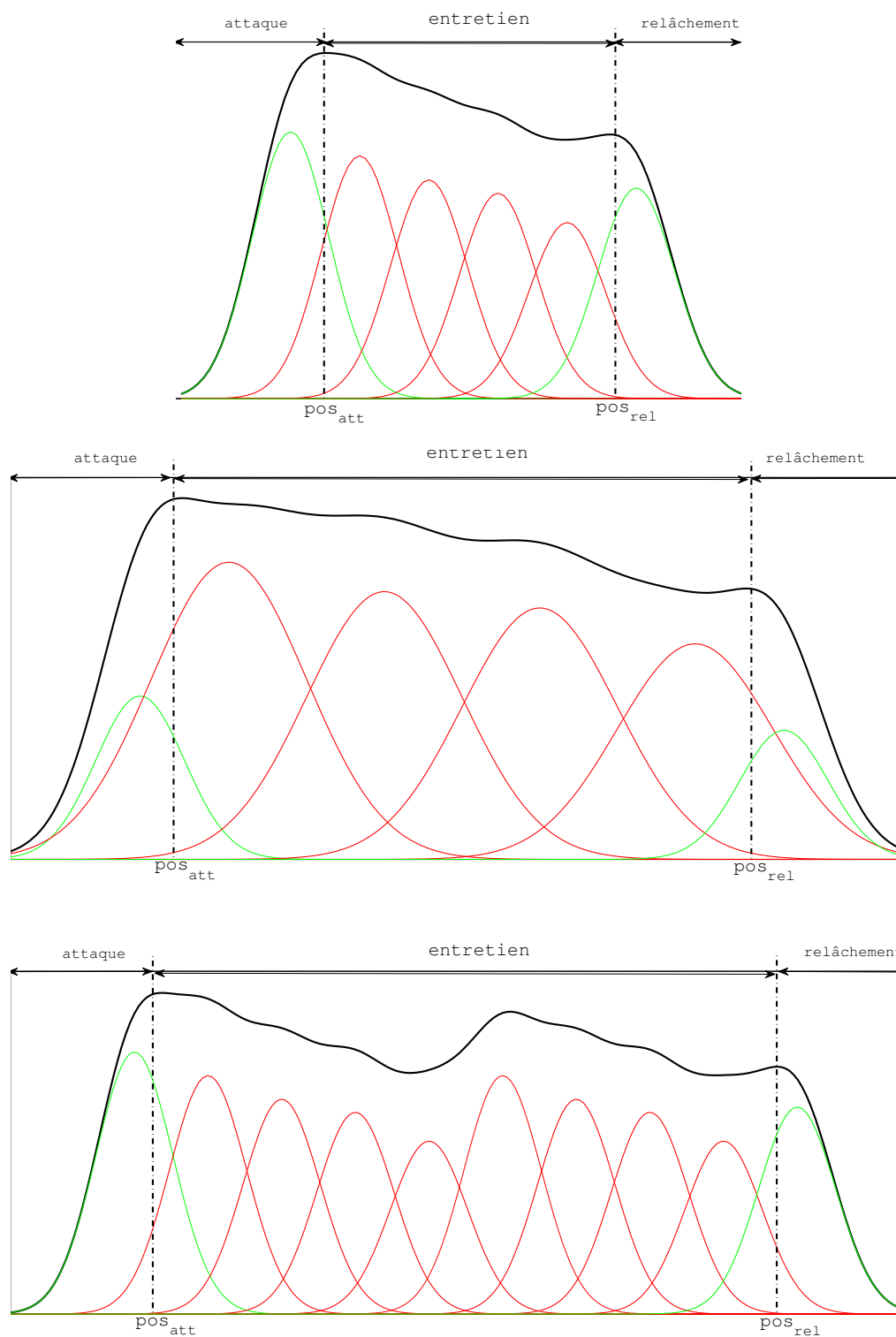


Figure 6.3: Effet d'une modification de durée pertinente par dilatation et comparaison avec la répétition moins pertinente vis-à-vis des données du modèle

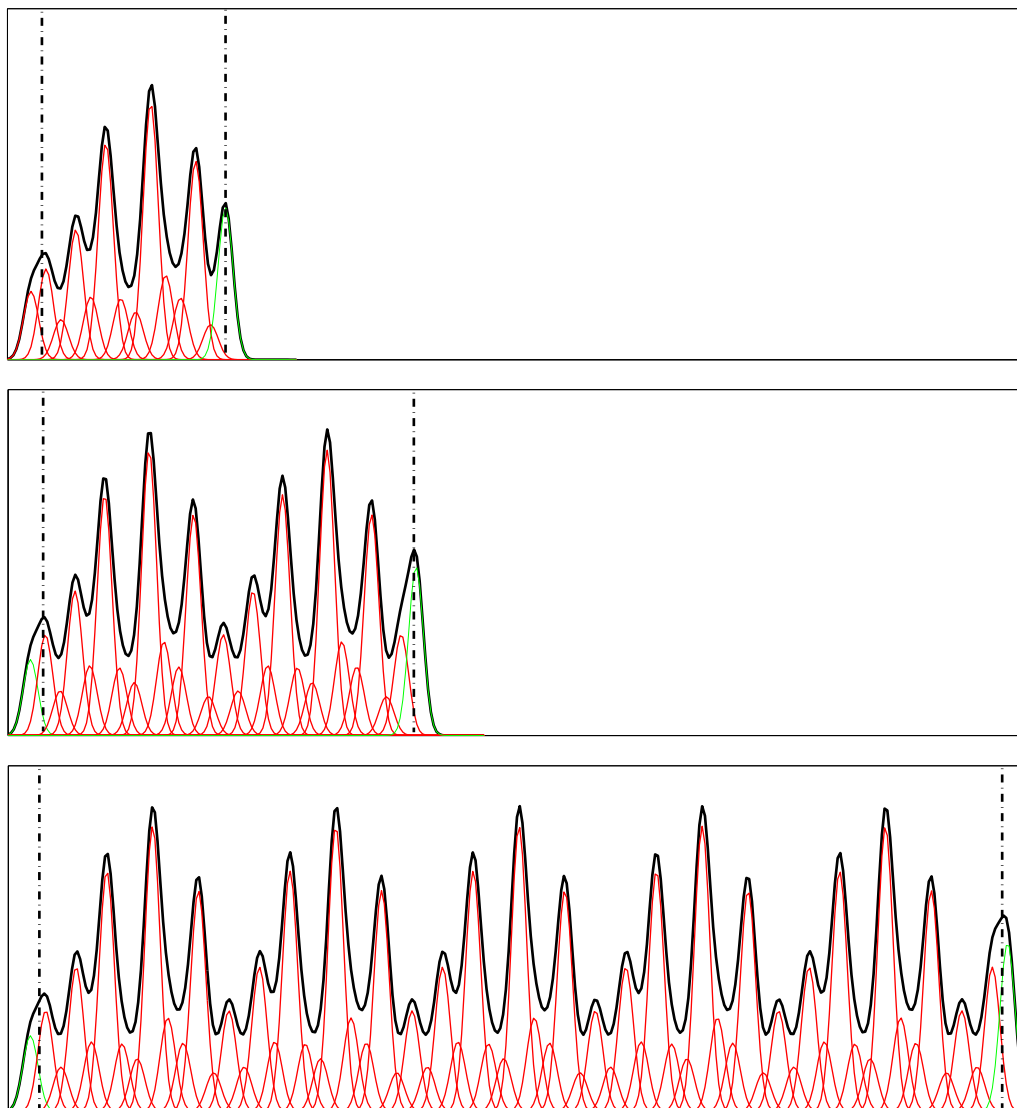


Figure 6.4: Effet d'une modification de durée pertinente par répétition (cas du tremolo) avec des facteurs 2 et 5

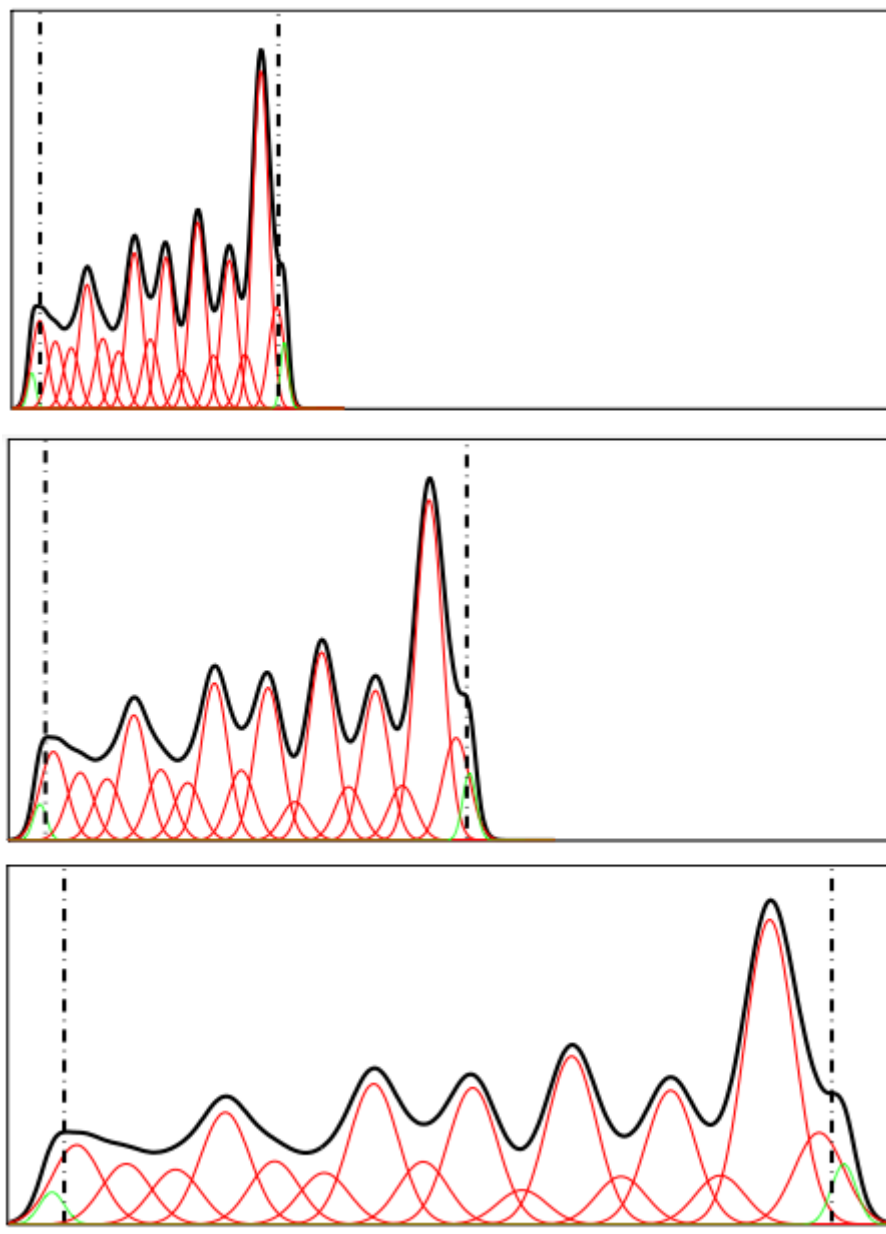


Figure 6.5: Effet d'une modification de durée pertinente par répétition (cas du crescendo) avec des facteurs 2 et 5

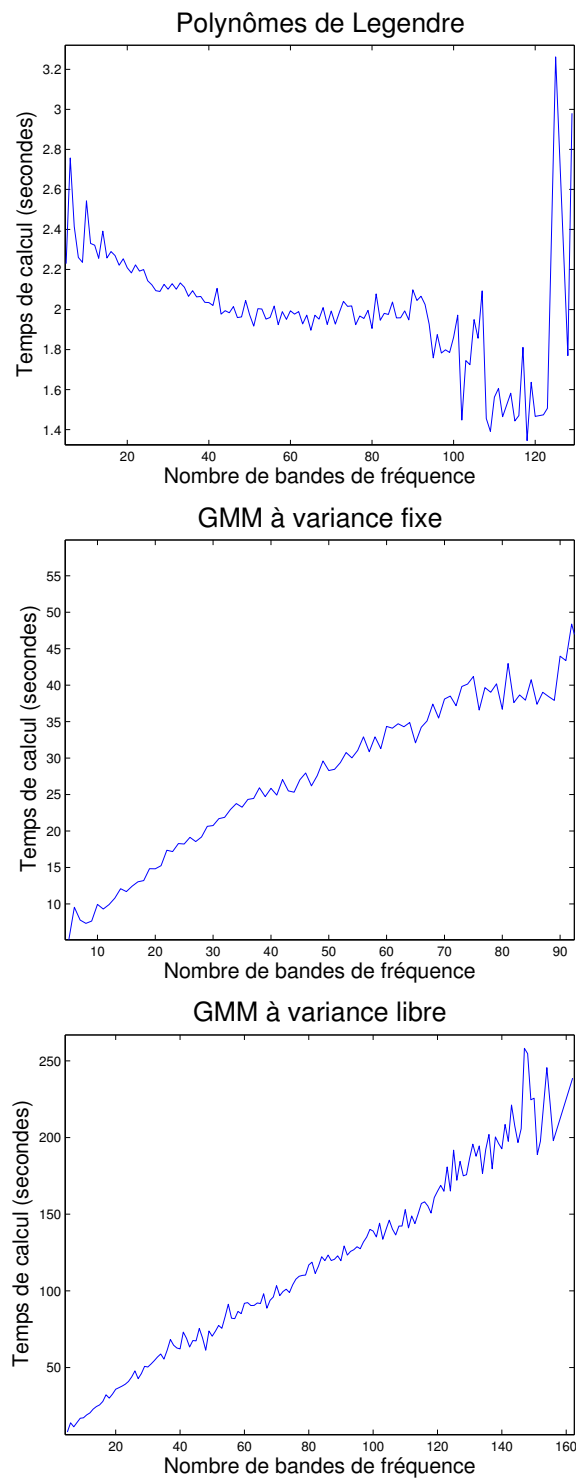


Figure 6.6: Comparaisons des temps de calcul moyens des différents modèles en fonction du nombre de bandes de fréquence

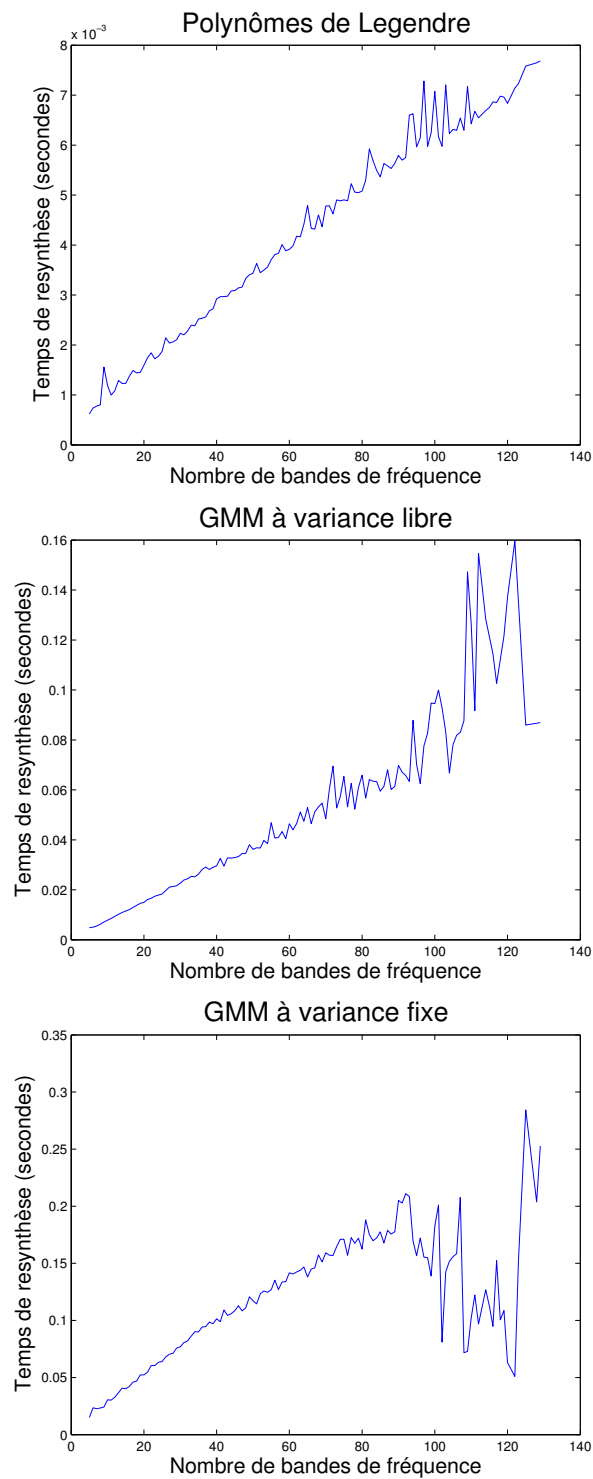


Figure 6.7: Comparaisons des temps de resynthèse moyens des différents modèles en fonction du nombre de bandes de fréquence

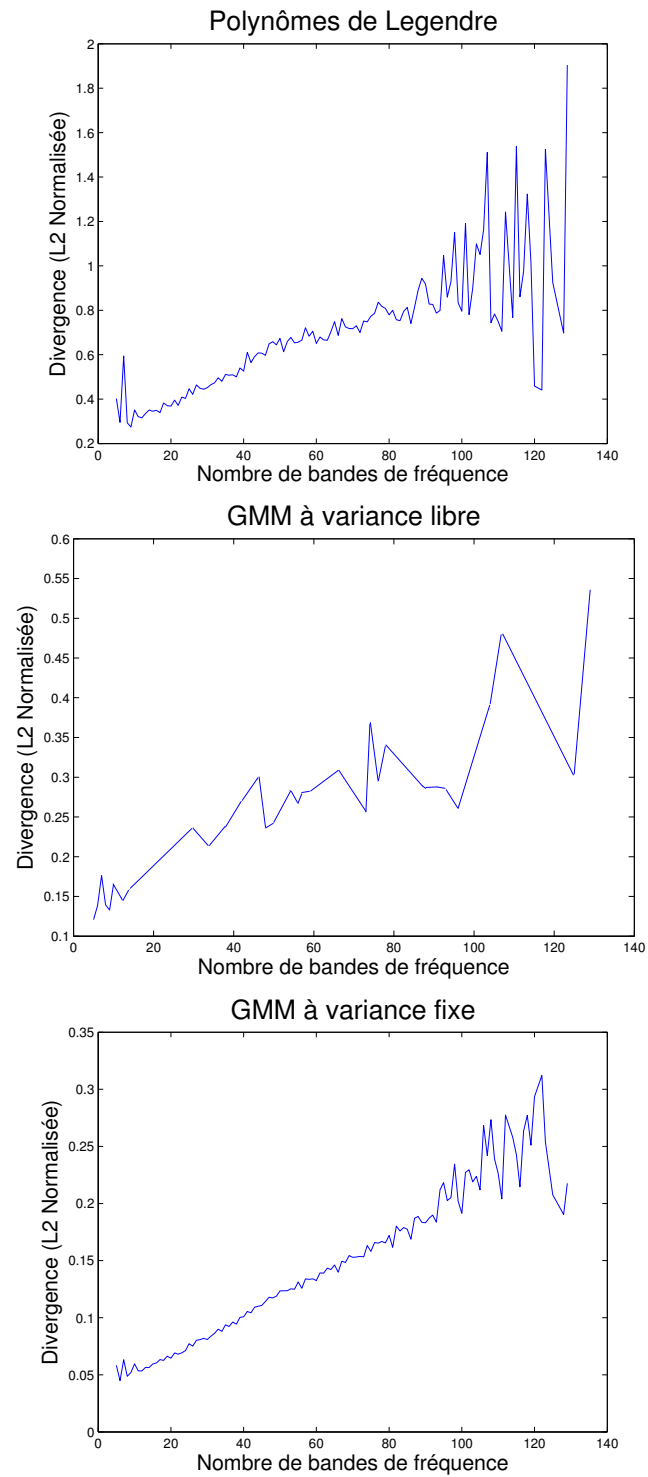


Figure 6.8: Précision moyenne des différents modèles en fonction du nombre de bandes de fréquence

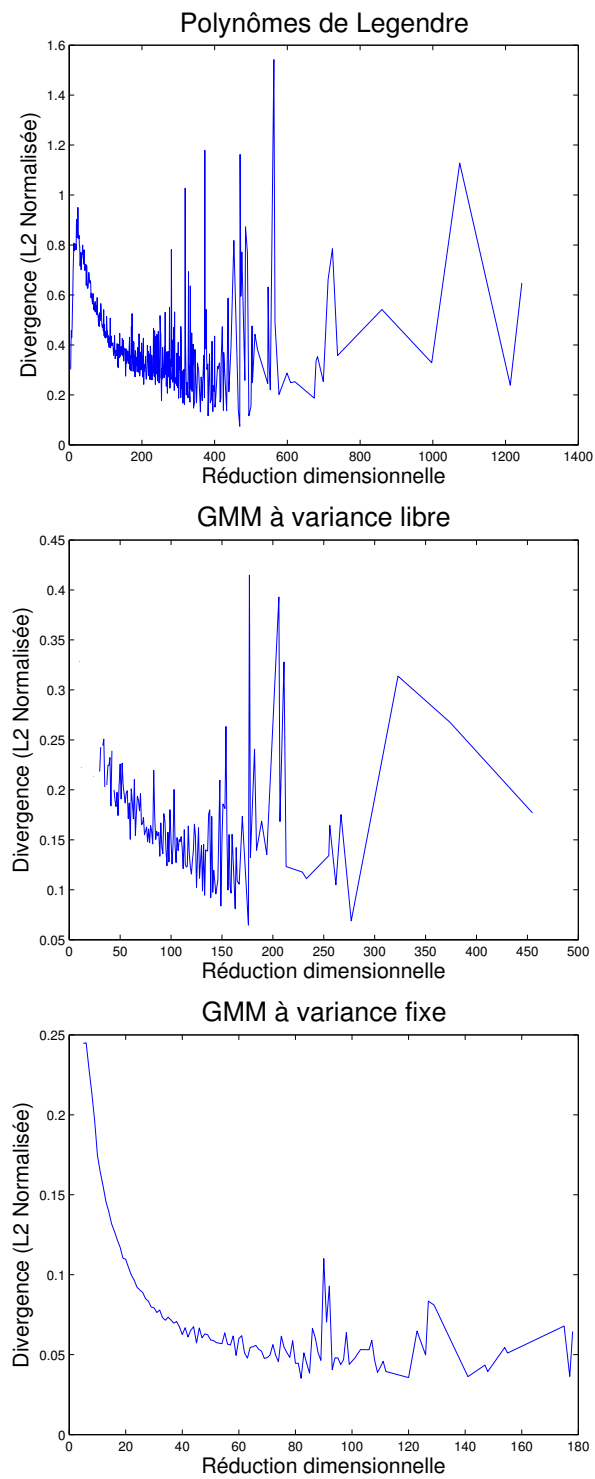


Figure 6.9: Précision moyenne des différents modèles en fonction de la réduction dimensionnelle induite

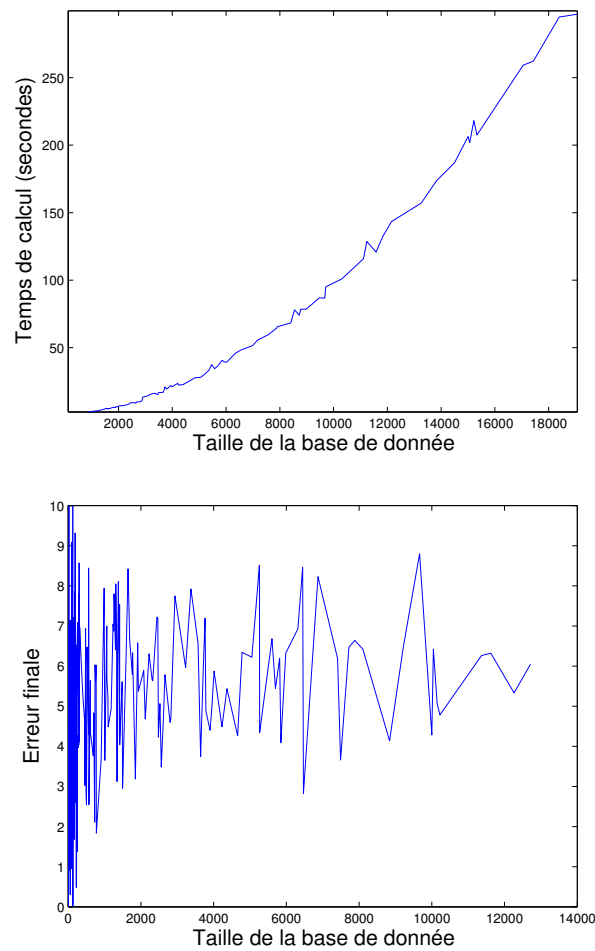


Figure 6.10: Temps de calcul moyens et erreur finale moyenne de l'algorithme en fonction de la taille de la base de données

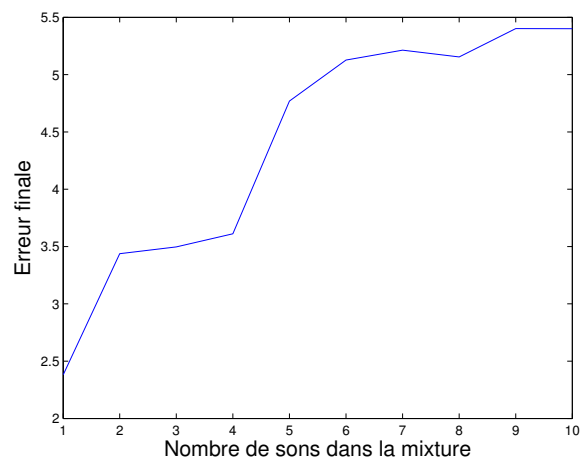


Figure 6.11: Erreur finale moyenne de l'algorithme en fonction du nombre de sons dans la mixture

Précision		Nombre de sons dans la mixture							
Instruments	1	2	3	4	5	6	7	8	9
1	0.997	0.982	0.943	0.971	0.920	0.907	0.862	0.881	0.879
2	0.941	0.925	0.914	0.923	0.902	0.912	0.915	0.892	0.863
3	0.843	0.832	0.881	0.912	0.845	0.861	0.805	0.842	0.833
4	0.741	0.781	0.794	0.765	0.762	0.705	0.731	0.714	0.720
5	0.689	0.704	0.692	0.673	0.662	0.652	0.671	0.641	0.538
6	0.646	0.687	0.661	0.654	0.664	0.602	0.614	0.634	0.510
7	0.607	0.612	0.587	0.594	0.619	0.581	0.567	0.531	0.592
8	0.581	0.551	0.541	0.512	0.566	0.574	0.617	0.522	0.558
9	0.554	0.572	0.564	0.524	0.533	0.554	0.537	0.518	0.513
10	0.510	0.501	0.517	0.490	0.475	0.501	0.466	0.481	0.492
11	0.438	0.422	0.443	0.515	0.445	0.423	0.391	0.399	0.412
12	0.349	0.407	0.358	0.346	0.297	0.323	0.317	0.275	0.307

Recall		Nombre de sons dans la mixture							
Instruments	1	2	3	4	5	6	7	8	9
1	1.000	0.978	0.836	0.878	0.736	0.708	0.689	0.641	0.613
2	1.000	0.914	0.786	0.763	0.749	0.720	0.619	0.525	0.584
3	1.000	0.875	0.756	0.737	0.712	0.681	0.658	0.548	0.608
4	1.000	0.853	0.734	0.718	0.705	0.698	0.629	0.579	0.572
5	1.000	0.788	0.779	0.720	0.655	0.712	0.674	0.562	0.582
6	1.000	0.752	0.799	0.712	0.716	0.683	0.614	0.607	0.567
7	1.000	0.771	0.751	0.708	0.699	0.751	0.665	0.575	0.573
8	1.000	0.761	0.769	0.708	0.659	0.653	0.639	0.688	0.651
9	1.000	0.783	0.685	0.771	0.628	0.683	0.628	0.587	0.622
10	1.000	0.774	0.653	0.682	0.677	0.647	0.587	0.570	0.605
11	1.000	0.761	0.785	0.775	0.627	0.612	0.585	0.601	0.555
12	1.000	0.769	0.643	0.660	0.680	0.611	0.582	0.637	0.542

F-Measure		Nombre de sons dans la mixture							
Instruments	1	2	3	4	5	6	7	8	9
1	0.999	0.980	0.886	0.922	0.818	0.795	0.766	0.742	0.722
2	0.970	0.920	0.845	0.835	0.818	0.805	0.738	0.661	0.697
3	0.915	0.853	0.814	0.815	0.772	0.761	0.724	0.664	0.703
4	0.851	0.815	0.763	0.741	0.732	0.702	0.676	0.640	0.638
5	0.816	0.743	0.733	0.696	0.659	0.681	0.673	0.599	0.559
6	0.785	0.718	0.724	0.682	0.689	0.640	0.614	0.620	0.537
7	0.755	0.682	0.659	0.646	0.657	0.655	0.612	0.552	0.582
8	0.735	0.640	0.635	0.594	0.609	0.611	0.628	0.594	0.601
9	0.713	0.661	0.619	0.624	0.577	0.612	0.579	0.550	0.562
10	0.676	0.608	0.577	0.570	0.558	0.565	0.520	0.522	0.543
11	0.610	0.543	0.566	0.620	0.520	0.500	0.469	0.480	0.473
12	0.517	0.532	0.460	0.454	0.413	0.423	0.410	0.384	0.392

Table 6.1: Précision, Recall et F-Measure par rapport au nombre d'instruments dans la base de données pour un facteur de parcimonie de 0.97

Précision		Nombre de sons dans la mixture							
Instruments	1	2	3	4	5	6	7	8	9
1	0.967	0.853	0.857	0.768	0.753	0.734	0.763	0.744	0.756
2	0.837	0.874	0.805	0.696	0.635	0.742	0.706	0.748	0.739
3	0.495	0.641	0.643	0.642	0.614	0.687	0.623	0.689	0.695
4	0.481	0.524	0.602	0.598	0.515	0.509	0.650	0.637	0.697
5	0.589	0.430	0.521	0.558	0.387	0.402	0.439	0.523	0.572
6	0.494	0.498	0.491	0.546	0.389	0.349	0.408	0.436	0.538
7	0.261	0.348	0.278	0.363	0.412	0.365	0.315	0.421	0.408
8	0.202	0.314	0.285	0.336	0.313	0.246	0.284	0.214	0.303
9	0.217	0.274	0.323	0.273	0.354	0.251	0.196	0.188	0.217
10	0.179	0.435	0.165	0.316	0.278	0.196	0.108	0.092	0.242
11	0.180	0.461	0.191	0.153	0.261	0.205	0.169	0.161	0.143
12	0.168	0.101	0.132	0.183	0.071	0.099	0.189	0.119	0.145

Recall		Nombre de sons dans la mixture							
Instruments	1	2	3	4	5	6	7	8	9
1	1.000	1.000	0.956	0.910	0.905	0.845	0.833	0.785	0.756
2	1.000	0.995	0.986	0.903	0.886	0.823	0.812	0.808	0.702
3	1.000	0.988	0.890	0.879	0.812	0.810	0.802	0.782	0.724
4	1.000	0.976	0.876	0.856	0.843	0.792	0.803	0.767	0.689
5	1.000	0.929	0.891	0.867	0.815	0.765	0.778	0.736	0.705
6	1.000	0.884	0.882	0.812	0.785	0.771	0.782	0.740	0.682
7	1.000	0.877	0.853	0.834	0.774	0.741	0.763	0.724	0.694
8	1.000	0.843	0.847	0.821	0.786	0.720	0.731	0.709	0.662
9	1.000	0.838	0.786	0.820	0.714	0.743	0.704	0.681	0.650
10	1.000	0.847	0.824	0.765	0.782	0.701	0.722	0.692	0.681
11	1.000	0.861	0.831	0.782	0.752	0.711	0.682	0.634	0.643
12	1.000	0.819	0.769	0.705	0.710	0.689	0.625	0.645	0.612

F-Measure		Nombre de sons dans la mixture							
Instruments	1	2	3	4	5	6	7	8	9
1	0.983	0.921	0.904	0.833	0.822	0.786	0.797	0.764	0.756
2	0.911	0.931	0.886	0.786	0.740	0.780	0.755	0.777	0.720
3	0.662	0.778	0.747	0.742	0.699	0.743	0.701	0.733	0.709
4	0.650	0.681	0.714	0.704	0.640	0.620	0.718	0.696	0.693
5	0.741	0.588	0.658	0.679	0.525	0.527	0.561	0.612	0.631
6	0.661	0.637	0.631	0.653	0.520	0.481	0.536	0.549	0.602
7	0.414	0.498	0.420	0.505	0.538	0.489	0.446	0.532	0.514
8	0.336	0.457	0.427	0.477	0.448	0.366	0.409	0.329	0.416
9	0.356	0.413	0.460	0.410	0.473	0.375	0.307	0.295	0.325
10	0.304	0.575	0.275	0.447	0.410	0.306	0.188	0.162	0.357
11	0.305	0.601	0.310	0.256	0.388	0.318	0.271	0.257	0.234
12	0.288	0.180	0.226	0.291	0.129	0.173	0.290	0.201	0.235

Table 6.2: Précision, Recall et F-Measure par rapport au nombre d'instruments dans la base de données pour un facteur de parcimonie de 0.7

List of Figures

3.1	Modélisation d'une enveloppe par mélange gaussien à variance libre	22
3.2	Modélisation d'une enveloppe par mélange gaussien à variance fixe	24
3.3	Deux approches de la modification de durée	25
3.4	Résultats de l'orchestration d'une voix extraite de <i>Poetry</i> de Marco Suarez en utilisant uniquement des descripteurs statiques	30
3.5	Résultats de l'orchestration de la même cible après intégration de notre modèle temporel	30
3.6	Comparaison des résultats de l'orchestration de la voix de <i>Chewbacca</i> en utilisant uniquement des descripteurs statiques (gauche) et après intégration du modèle temporel (droite)	31
4.1	Utilisation de l'algorithme NMF (image tirée de [42]) pour la reconnaissance d'images faciales.	35
4.2	Un vecteur synthétique (en haut à gauche) et les résultats d'une projection parcimonieuse pour des valeurs $l_\epsilon = 0.4, 0.7, 0.9$	38
4.3	Comparaison entre l'approche alternée (gauche) et l'algorithme original (droite) en termes d'évolution de la fonction de divergence et donc de convergence	41
6.1	Comparaison entre transformée de Fourier et transformée à Q constant pour une note de piano (haut) et une séquence de sitar (bas)	49
6.2	Effet d'une modification de durée pertinente par répétition et comparaison avec la dilatation moins pertinente vis-à-vis des données du modèle	50
6.3	Effet d'une modification de durée pertinente par dilatation et comparaison avec la répétition moins pertinente vis-à-vis des données du modèle	51
6.4	Effet d'une modification de durée pertinente par répétition (cas du tremolo) avec des facteurs 2 et 5	52
6.5	Effet d'une modification de durée pertinente par répétition (cas du crescendo) avec des facteurs 2 et 5	53
6.6	Comparaisons des temps de calcul moyens des différents modèles en fonction du nombre de bandes de fréquence	54
6.7	Comparaisons des temps de resynthèse moyens des différents modèles en fonction du nombre de bandes de fréquence	55
6.8	Précision moyenne des différents modèles en fonction du nombre de bandes de fréquence	56
6.9	Précision moyenne des différents modèles en fonction de la réduction dimensionnelle induite	57
6.10	Temps de calcul moyens et erreur finale moyenne de l'algorithme en fonction de la taille de la base de données	58
6.11	Erreur finale moyenne de l'algorithme en fonction du nombre de sons dans la mixture	58

List of Tables

3.1	Résultats de discrimination pour différentes simplifications spectro-temporelles	17
3.2	Base de données utilisée pour l'analyse	26
3.3	Base de données à forte variation temporelle	27
3.4	Temps de resynthèse, réduction dimensionnelle et précision des modèles	28
3.5	Temps de calcul totaux par famille	28
3.6	Temps de calcul moyens et variances en secondes	28
6.1	Précision, Recall et F-Measure par rapport au nombre d'instruments dans la base de données pour un facteur de parcimonie de 0.97	59
6.2	Précision, Recall et F-Measure par rapport au nombre d'instruments dans la base de données pour un facteur de parcimonie de 0.7	60

Bibliography

- [1] Abe, T. Itoyama, K. Yoshii, K. Komatani, K. Ogata, T. and Okuno H.G. "Analysis-and-manipulation approach to pitch and duration of musical instrument sounds without distorting timbral characteristics", *Proceedings of the 11th Conference on Digital Audio Effects*, pp. 249-256, Espoo, Finland, 2008.
- [2] Abdallah, S.A. and Plumbley, M.D. "Polyphonic transcription by non-negative sparse coding of power spectra", *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*, pp 318-325, Barcelona, Spain, October 10-14, 2004.
- [3] Atlas, L. Janssen, C. "Coherent modulation spectral filtering for single-channel music source separation", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 461-464, Philadelphia, PA, USA, 2005.
- [4] Ballet, G. Borghesi, R. Hoffmann, P. and Lévy, F. "Studio Online 3.0 : An Internet "Killer Application" for Remote Access to IRCAM Sounds and Processing tools", *Actes des Journées d'informatique musicale*, Paris, France, 1999.
- [5] Benetos E., Kotti M. and Kotropoulos C. "Applying supervised classifiers based on non-negative matrix factorization to musical instrument classification", *Proceedings of the IEEE International Conference on Multimedia*, pp. 2105-2108, Toronto, Canada, 2006.
- [6] Berlioz, H. *Traité d'instrumentation et d'orchestration*, Henri Lemoine ed., Paris, 1855.
- [7] Bertin, N. Badeau, R. and Richard, G. "Blind signal decompositions for automatic transcription of polyphonic music : NMF and K-SVD on the benchmark", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 65-68, Honolulu, Hawaii, 2007.
- [8] Bertin, N. Fevotte, C. and Badeau R. "A tempering approach for itakura-saito non-negative matrix factorization with application to music transcription", *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing ICASSP'09*, Taipei, Taiwan, 2009.
- [9] Bregman, A.S. "Timbre, orchestration, dissonance, et l'organisation auditive", *Le timbre, métaphores pour la composition*, Ed. J.-B. Barrière, Paris, 1991.
- [10] Brown, J.C. "Calculation of a constant Q spectral transform", *Journal of the Acoustic Society of America*, vol. 90, pp. 60-66, 1991.
- [11] Brown, J.C. and Puckette, M.S. "An efficient algorithm for the calculation of a constant Q transform", *Journal of the Acoustic Society of America*, vol. 92, no.5, pp. 2698-2701, 1992.
- [12] Carpentier, G. Tardieu, D. Assayag, G. Rodet, X. and Saint-James, E. "Imitative and generative orchestrations using pre-analyzed sound databases", *Proceedings of the 3rd Sound and Music Computing Conference*, pp. 115-122, Marseille, France, 2006.
- [13] Carpentier, G. Tardieu, D. Assayag, G. Rodet, X. and Saint-James, E. "An evolutionary approach to computer-aided orchestration", *Lecture Notes in Computer Science : Applications of Evolutionary Computing*, vol. 4448, pp. 488-497, Springer, 2007.

- [14] Carpentier, G. "Approche computationnelle de l'orchestration musicale - Optimisation multicritere sous contraintes de combinaisons instrumentales dans de grandes banques de sons", Ph.D. Thesis, University UPMC Paris 6, 2008.
- [15] Casella, A. *Technique de l'orchestre contemporain*, Ricordi, Paris, 1958.
- [16] Cont, A. "Realtime audio to score alignment for polyphonic music instruments using sparse non-negative constraints and Hierarchical HMMs", *IEEE International Conference in Acoustics and Speech Signal Processing (ICASSP)*, May 2006.
- [17] Cont, A. "Realtime multiple pitch observation using sparse non-negative constraints.", *International Symposium on Music Information Retrieval (ISMIR)*, October 2006.
- [18] Cont, A. Dubnov, S. and Wessel, D. "Realtime multiple-pitch and multiple-instrument recognition for music signals using sparse non-negative constraints", *Proceedings of the 10th Conference on Digital Audio Effects*, Bordeaux, France, 2007.
- [19] Dubnov, S. and Rodet, X. "Investigation of phase coupling phenomena in sustained portion of musical instruments sound", *Journal of the Acoustical Society of America*, vol. 113, pp. 348–359, 2003.
- [20] Duda, R.O. Hart, P.E. and Stork, D.G. *Pattern Classification*, John Wiley & Sons, 2nd edition, 2001
- [21] Erickson, R. *Sound structure in music*, University of California Press, 1975.
- [22] Everitt, B.S. and Hand, D.J. Finite Mixture Distributions, *Monographs on Applied Probability and Statistics*, Chapman and Hall, 1981.
- [23] Faure, A. "Des sons aux mots : Comment parle-t-on du timbre musical ?", Ph.D. Thesis, Ecoles des Hautes Etudes en Sciences Sociales, Sciences Cognitives, Paris, 2000.
- [24] Figueiredo, M.A.T. and Jain, A.K. "Unsupervised learning of finite mixture models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, 2002.
- [25] Finesso, L. and Spreij, P. "Approximate nonnegative matrix factorization via alternating minimization", *Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems*, Leuven, Belgium, 2004.
- [26] FitzGerald, D. Cranitch, M. and Coyle, E. "Non-negative tensor factorisation for sound source separation", *Proceedings of the Irish Signals and Systems Conference*, pp. 8–12, Dublin, Ireland, 2005.
- [27] FitzGerald, D. Cranitch, M. and Coyle, E. "Shifted non-negative matrix factorization for sound source separation", *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse, France, 2006.
- [28] Goto, M. Hashiguchi, H. Nishimur, T. and Oka, R. "RWC music database : music genre database and musical instrument sound database", *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pp. 229–230, 2003.
- [29] Goto, M. "Development of the RWC music database", *Proceedings of the 18th International Congress on Acoustics (ICA 2004)*, vol. 1, pp. 553–556, 2004.
- [30] Guillaumet, D. Bressan, M. and Vitria, J. "A weighted non-negative matrix factorization for local representations", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 942–947, 2001.

- [31] Guillaumet, D. Vitrià, J. and Schiele, B. "Introducing a weighted non-negative matrix factorization for image classification", *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2447-2454, 2003.
- [32] Helén, M. and Virtanen, T. "Perceptually motivated parametric representation for harmonic sounds for data compression purposes", *Proceedings of the 6th International Conference on Digital Audio Effects*, London, UK, 2003.
- [33] Helén, M. and Virtanen, T. "Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine", *Proceedings of the 13th European Signal Processing Conference*, Antalya, Turkey, 2005.
- [34] Helmholtz, H. *On the sensations of tone as a physiological basis for the theory of music*. 1863, Dover Publications Inc., New York, 1954.
- [35] Hoyer, P.O. "Non-negative matrix factorization with sparseness constraints", *Journal of Machine Learning Research*, vol. 5, pp. 1457-1469, 2004.
- [36] Hummel, T.A. "Simulation of human voice timbre by orchestration of acoustic music instruments", *Proceedings of the International Computer Music Conference*, Barcelona, Spain, 2005.
- [37] Koechlin, C. *Traité de l'orchestration*, Max Eschig, Paris, 1943.
- [38] Itoyama, K. Goto, M. Komatani, K. Ogata, T. and Okuno H.G. "Integration and adaptation of harmonic and inharmonic models for separating polyphonic musical signals", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, Hawaii, USA, 2007.
- [39] Jensen, K. "Musical Instruments Parametric Evolution", *Proceedings of the ISMA*, Mexico City, Mexico, 2002.
- [40] Kameoka, H. Nishimoto, T. and Sagayama, S. "Harmonic-temporal-structured clustering via deterministic annealing EM algorithm for audio feature extraction", *Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR)*, London, England, 2005.
- [41] Kameoka, H. Nishimoto, T. and Sagayama, S. "A multipitch analyzer based on harmonic temporal structured clustering", *IEEE Transactions on Audio, Speech and language processing*, vol. 15, no. 3, 2007.
- [42] Lee, D. and Seung, H. "Learning the parts of objects by non-negative matrix factorization", *Nature*, vol. 401, pp. 788-791, 1999.
- [43] Lee, D. and Seung, H. "Algorithms for non-negative matrix factorization", *Proceedings of the Conference on Neural Information Processing Systems*, pp. 556-562, Vancouver, Canada, 2001.
- [44] Le Roux, J. Kameoka, H. Ono, N. De Chevigné, A. and Sagayama, S. "Harmonic-temporal clustering of speech for single and multiple f_0 contour estimation in noisy environments", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, Hawaii, USA, 2007.
- [45] Lin, C.-J. "On the convergence of multiplicative update algorithms for nonnegative matrix factorization", *IEEE Transactions on Neural Networks*, vol. 18, n. 6, pp. 1589-1596, 2007.
- [46] McAdams, S. Beauchamp, J.W. and Meneguzzi, S. "Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters", *Journal of the Acoustical Society of America*, vol. 105, no. 2, pp. 882-897, 1999.
- [47] Mel, B.W. "Computational neuroscience : Think positive to find parts", *Nature*, vol. 401, pp. 759-760, 1999.

- [48] Miyamoto, K. Kameoka, H. Takeda, H. Nishimoto, T. and Sagayama, S. “Probabilistic approach to automatic music transcription from audio signals”, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, Hawaii, USA, 2007.
- [49] Miyamoto, K. Kameoka, H. Nishimoto, T. Ono, N. and Sagayama, S. “Harmonic-Temporal-Timbral Clustering (HTTC) for the analysis of multi-instrument polyphonic music signals”, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, Nevada, USA, 2008.
- [50] Paatero, P. and Tapper, U. “Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values”, *Environmetrics*, vol. 5, pp. 111–126, 1994.
- [51] Paulus, J. and Virtanen, T., “Drum transcription with non-negative spectrogram factorisation”, *Proceedings of the 13th European Signal Processing Conference*, Antalya, Turkey, 2005.
- [52] Peeters, G. “Automatic classification of large musical instrument databases using hierarchical classifiers with inertia ratio maximization”, *Proceedings of the 115th Audio Engineering Society Convention*, New York, USA 2003.
- [53] Psenicka, D. “Sporch: An algorithm for orchestration based on spectral analyses of recorded sounds”, *Proceedings of International Computer Music Conference*, Singapore, 2003.
- [54] Piston, W. *Orchestration*, Norton Company, New York, 1955.
- [55] Raczynski, S.A. Ono, N. and Sagayama, S. “Multipitch analysis with harmonic nonnegative matrix approximation”, *Proceedings of the 8th International Conference on Music Information Retrieval*, Vienna, Austria, 2007.
- [56] Rimski-Korsakov, N.A. *Principles of Orchestration, with Musical Examples Drawn from his own Works*, Maximilian Steinberg, New York, 1964.
- [57] Risset, J.C. “Computer study of trumpet tones”, *Journal of the Acoustical Society of America*, vol. 33, 1965.
- [58] Rose, F. and Hetrick, J. “Spectral analysis as a resource for contemporary orchestration technique”, *Proceedings of Conference on Interdisciplinary Musicology*, Montreal, Canada, 2005.
- [59] Schaffer, P. *Traité des objets musicaux*, Seuil, Paris, 1966.
- [60] Schoenberg, A. *Harmonielehre*, Universal Edition, Vienne, 1911.
- [61] Schmidt, M.N. and Olsson, R.K. “Single-Channel Speech Separation using Sparse Non-Negative Matrix Factorization”, *Proceedings of the Ninth International Conference on Spoken Language Processing*, Pittsburgh, Pennsylvania, 2006.
- [62] Sha, F. and Saul, L.K. “Realtime pitch determination of one or more voices by non-negative matrix factorization”, *Proceedings of the 17th Conference of Advances in Neural Information Processing Systems*, Vancouver, Canada, 2005.
- [63] Strawn, J. “Approximation and syntactic analysis of amplitude and frequency functions for digital sound synthesis”, *Computer Music Journal*, vol. 4, no. 3, pp. 3-24, 1980.
- [64] Sukittanon, S. Atlas, L.E. and Pitton, J.W. “Modulation-scale analysis for content identification”, *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 3023– 3035, 2004.
- [65] Tardieu, D. Peeters, G. and Rodet, X. “An instrument timbre model for computer aided orchestration”, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, 2007.

- [66] Tardieu, D. Carpentier, G. and Rodet, X. "Computer-aided orchestration based on probabilistic instruments models and genetic exploration", *Proceedings of International Computer Music Conference*, Copenhagen, Denmark, 2007.
- [67] Tardieu, D. Carpentier, G. Peeters, G. "Instrument Sound Description and Modelization in the Context of Computer Aided Orchestration", *Proceedings of Acoustics'08*, Paris, France, 2008.
- [68] Tardieu, D. "Modèles d'instruments pour l'aide à l'orchestration" Ph.D. Thesis, University UPMC Paris 6, 2008.
- [69] Theodoridis, S. and Koutroumbas, K. *Pattern Recognition*, Academic Press, 1999.
- [70] Tisserand, P. "ECRINS : Rapport sur l'évolution dynamique", Rapport interne, IRCAM, 2002.
- [71] Travis, F.I. *Verdi's orchestration*, ed. Juris-Verlag, Zürich, 1956.
- [72] Vaggione, H. "Timbre as syntax : A spectral modeling approach", *Contemporary Music Review*, vol. 10, no. 2, pp. 73-83, 1994.
- [73] Verbeek, J.J. Vlassis, N. and Kröse, B. "Efficient greedy learning of Gaussian mixture models" *Neural Computation*, vol. 5, no. 2, pp. 469-485, 2003.
- [74] Vincent, E. Bertin, N. and Badeau, R. "Two nonnegative matrix factorization methods for polyphonic pitch transcription", *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienne, Septembre 2007.
- [75] Vincent, E. Bertin, N. and Badeau, R. "Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription", *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing ICASSP'08*, pp. 109-112, Las Vegas, Nevada, 2008.
- [76] Virtanen, T. and Klapuri, A. "Analysis of polyphonic audio using source-filter model and non-negative matrix factorization", *Advances in Models for Acoustic Processing, Neural Information Processing Systems Workshop*, 2006.
- [77] Virtanen, T. "Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria", *IEEE Transactions on Audio, Speech, and Language Processing*, vol 15, no. 3, March 2007.
- [78] Virtanen, T. Mesaros, A. and Ryyänen, M. "Combining pitch-based inference and non-negative spectrogram factorization in separating vocals from polyphonic music", *ISCA Tutorial and Research Workshop on Statistical And Perceptual Audition*, Brisbane, Australie, 2008.
- [79] *Vienna Symphonic Library*, <http://www.vsl.co.at>.
- [80] Wang, B. and Plumbley, M.D. "Musical audio stream separation by non-negative matrix factorization", *Proceedings of the DMRN Summer Conference*, Glasgow, Scotland, UK, 2005.
- [81] Wang, B. and Plumbley, M.D. "Investigating single-channel audio source separation methods based on non-negative matrix factorization", *Proceedings of the ICA Research Network International Workshop*, Liverpool, UK, pp 17-20, 2006.
- [82] Welburn, S. J. Plumbley, M.D. and Vincent, E. "Object-coding for resolution-free musical audio", *Proceedings of the 31st International AES Conference: "New Directions in High Resolution Audio"*, London, UK, 2007.
- [83] Wilson, K.W. Raj, B. Smaragdis, P. and Divakaran, A. "Speech denoising using non-negative matrix factorization with priors", *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4029-4032, 2008.