



HAL
open science

Composer les mélanges sonores avec les maquettes spectrales

Philippe Esling, Carlos Agon

► **To cite this version:**

Philippe Esling, Carlos Agon. Composer les mélanges sonores avec les maquettes spectrales. Journées d'Informatique Musicale, May 2010, Rennes, France. hal-01106581

HAL Id: hal-01106581

<https://hal.science/hal-01106581>

Submitted on 10 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

COMPOSER LES MÉLANGES SONORES AVEC LES MAQUETTES SPECTRALES

Philippe ESLING Carlos AGON

Institut de Recherche et Coordination Acoustique / Musique
1, place Igor Stravinsky - 75004 Paris
{esling, agonc}@ircam.fr

RÉSUMÉ

Nous présentons dans cet article un système novateur permettant d'explorer l'univers de mélanges sonores à des fins compositionnelles. Ce système intègre plusieurs paradigmes de composition et permet un lien entre la micro-structure du timbre et la macro-forme du discours musical, en imposant des relations temporelles sur des descripteurs spectraux. Il a également été conçu dans le but d'intégrer des unités de signal et de symbolique dans un cadre commun. Le système repose sur une architecture SQL qui agit comme un lexique organisé. Cette base de données contient à la fois des informations symboliques ainsi que des descripteurs spectraux calculés sur des sons provenant de plusieurs collections. Chaque descripteur est modélisé temporellement en utilisant un modèle de mélange gaussien à variance fixe (GMMFV). Nous introduisons également un système de requêtes temporelles permettant de trouver intuitivement des sons correspondants à des courbes d'évolution temporelle des descripteurs spectraux. Finalement, nous présentons les maquettes spectrales pour OpenMusic (OM), qui permettent des interactions temporelles et structurelles complexes pouvant être imposées entre des unités musicales de natures diverses.

1. INTRODUCTION

Au cours du siècle dernier, s'est opéré un revirement dans la musique instrumentale, tendant à dépasser de plus en plus les catégories traditionnelles de l'écriture (hauteur, durée et intensité) et s'attachant à l'inharmonique, la composition de sons bruités et à forte variation de timbre dans le temps. Les marques les plus évidentes de cette évolution sont la *musique spectrale*, la *musique concrète*, ainsi que les écoles de la complexité (Ferneyhough, Dillon) qui cherchent à saturer l'écriture musicale. Parmi les compositeurs contemporains, beaucoup sont résolument tournés vers les qualités expressives du son, grâce au potentiel des sons complexes, bruités ou issus de sources électroniques. L'avènement de la technologie a repoussé la frontière entre le bruit et la musique, comme le montre la démarche de Pierre Schaeffer. En effet, il estime [17] la nécessité de « remplacer la variété limitée de timbres instrumentaux possédée par un orchestre avec l'infnie variété de timbres bruités obtenus grâce à des mécanismes spéciaux ».

Depuis le XIX^e siècle, le timbre auparavant considéré comme une propriété intrinsèque des instruments s'en est progressivement détaché afin de devenir d'abord un élément nouveau dans le langage musical, évoluant jusqu'au point d'être considéré comme l'épine dorsale de l'écriture musicale. Arriva alors le temps de créer des timbres nouveaux, « composer le timbre » en puisant dans les multiples richesses du vocabulaire instrumental et électronique. Le timbre est devenu le *matériau* principal de la composition, organisant un ordre complexe qui s'élève au-dessus du système instrumental. Le langage musical s'est ainsi progressivement enrichi et ces innovations font maintenant partie intégrante du discours musical contemporain. Ainsi, considéré en termes d'écriture musicale, le timbre joue un double rôle suivant les deux pôles de l'univers instrumental que sont l'articulation et la fusion [5]. L'articulation du discours musical est basé sur la fonction identitaire du timbre, la fusion (ou illusion) qui assure l'impossibilité d'identification est utilisée pour orienter l'auditeur vers de nouveaux éléments du discours musical. Nous pouvons voir ici que chaque pôle s'appuie sur la notion de mélange sonore. Ainsi l'aventure des sons nouveaux, de la musique expérimentale, suscitée par les possibilités de synthèse et de traitement du son ont véritablement fait émerger le problème du timbre des mélanges sonores comme un axe central de recherche en informatique musicale. Dans cette ère nouvelle, l'écoute semble plus orientée sur le *son* de la *musique*. Il s'agit alors d'élever le signal au rang de musique. Ces nouvelles démarches appellent à un approfondissement de nos connaissances sur les mélanges sonores pour arriver à suivre et accompagner les nouvelles orientations musicales. Dans celles-ci, le timbre peut être considéré comme un vocabulaire d'expressivité. Pourtant, il a toujours semblé échapper à l'abstraction pure. Bien sûr, différents modèles de synthèse permettent aux compositeurs de produire des timbres. Avec la synthèse additive, par exemple, les compositeurs peuvent produire des sons par la superposition de plusieurs sinus qui sont considérés comme les unités primitives du système. Cependant, il semble impératif que ces unités primitives soient plus complexes et musicalement pertinentes, comme des sons instrumentaux ou synthétisés. Il est donc nécessaire d'imaginer un système aussi bien qualitatif que quantitatif afin d'explorer les possibilités d'expressivité du timbre. Une approche exploratoire permettrait d'établir

de nouvelles stratégies compositionnelles, profilant ainsi une nouvelle dimension dans l'imaginaire du son.

Cependant, nous pouvons déplorer l'absence de système de composition qui pourrait faire face à ces nouvelles pratiques. Le matériau musical a souvent été pensé et écrit pour un instrument. Aujourd'hui, les systèmes de composition suivent généralement ce paradigme traditionnel et sont inscrits dans un moment ponctuel de l'écriture. Cependant, de nombreux contextes musicaux (figures, textures, gestes) semblent relever conjointement de l'écriture et des mélanges sonores. La vision traditionnelle sur les mélanges sonores, à savoir l'orchestration, peut être vue comme l'ensemble des techniques d'écriture dans lequel le timbre agit comme paramètre principal. Aussi, au-delà du point de vue conventionnel, nous pouvons envisager une généralisation de cette question en comment les sons se combinent pour produire une unité musicale agréable. Dans cette perspective, les compositeurs superposent les unités élémentaires pour en former de plus grandes. Toutefois, dans leur pratique quotidienne de l'écriture, les compositeurs ne travaillent pas seulement par la superposition de sons stationnaires, mais plutôt par la sculpture, simultanément verticale et horizontale du son. Ainsi le processus du mélange des sons est directement lié à l'écriture, voire indissociable de la composition. Toutefois, le problème de l'écriture musicale est un domaine sensible pour la recherche scientifique étant donné que son évaluation technique est soumise à l'appréciation esthétique, et dont les mécanismes semblent difficiles à formaliser. Les fonctions d'analyse acoustique et de traitement symbolique des objets musicaux devraient donc coexister et interagir dans un environnement commun, subordonné à la pensée créatrice du compositeur.

Il semble que le processus compositionnel ne puisse pas contourner la notion de mélanges sonores. Toutefois, celle-ci représente un problème difficile qui soulève de nombreuses questions complexes. En effet, outre les possibilités infinies offertes par un seul signal, les mélanges sonores y ajoutent une nouvelle dimension combinatoire. Le premier problème est donc de trouver des unités sonores qui correspondent à des critères complexes et de natures variées. Il est possible de pallier cette question en mettant en place une base de données organisée. Cependant, celle-ci introduit les questions de taille de stockage ainsi que de recherche rapide basée à la fois sur des informations symboliques et spectrales. Le deuxième problème est la question fondamentale du mélange des sons et donc de trouver des opérations significatives à appliquer aux unités.

Afin de maîtriser la complexité des mélanges sonores, nous introduisons un système de composition qui offre des capacités d'exploration de cet univers combinatoire. L'objectif de ce système est d'abord de permettre un lien entre la micro-structure du timbre et la macro-forme du discours musical, en permettant d'imposer des relations temporelles à partir de descripteurs spectraux. Toutefois, il a également été conçu dans le but de permettre l'interaction entre des unités issues du signal et de la symbo-

lique dans un cadre commun. Afin de permettre des requêtes efficaces fondées sur des critères à la fois symboliques et spectraux, nous avons développé une architecture SQL spécifique. Cette base de données contient des informations symboliques sur les sons provenant de multiples collections ainsi qu'un ensemble complet de descripteurs spectraux pré-calculés sur ces sons. Comme nous l'avons dit précédemment, nous voulons utiliser la micro-structure du timbre comme un élément clé dans le processus compositionnel. Afin de préserver toute la structure spectro-temporelle, chaque descripteur est stocké dans la base de données en utilisant une modélisation temporelle. Les avantages de cette approche sont doubles. Premièrement, la modélisation permet une forte réduction des données en lissant la forme des descripteurs. Deuxièmement, en gardant la structure temporelle du timbre, il est également possible de faire des requêtes basées sur l'évolution des descripteurs plutôt que simplement sur des critères statiques à partir de descripteurs moyennés. Ainsi l'utilisateur peut trouver de manière intuitive des sons suivant une évolution déterminée de volume ou de fréquence fondamentale et ce tout en imposant des contraintes symboliques. Cette architecture permet donc d'établir un lexique organisé pour notre système. Enfin, après avoir permis la possibilité de recherches complexes sur les sons, nous présentons un système permettant d'utiliser pleinement ces nouvelles possibilités. Afin de composer avec des sons jusqu'au niveau du discours musical, nous utilisons le système de maquettes d'OpenMusic [3] comme point de départ. Une maquette peut être considérée comme une « métapartition » englobant dans un seul document, la notation musicale et des programmes audiovisuels [1]. Nous présentons dans cet article les maquettes spectrales, une extension des maquettes classiques permettant l'organisation dans le temps d'événements allant du macro-niveau du discours musical jusqu'aux propriétés micro-temporelles et spectrales.

Le reste de cet article est organisé comme suit. Comme la composition des mélanges sonores se situe à la croisée des mondes du signal et de la notation symbolique, nous discutons dans la section 2 des interactions entre ces deux domaines. Nous présentons ensuite dans la section 3 la modélisation temporelle qui permet d'utiliser les propriétés micro-structurelles du timbre comme un élément compositionnel central. Dans la section 4, nous introduisons la base de données SQL qui permet une organisation hiérarchique ainsi que des requêtes rapides basées sur des critères à la fois spectraux et symboliques. Enfin, dans la section 5 nous présentons le système des maquettes spectrales dans lequel tous ces éléments interagissent et comment des interactions temporelles et structurelles complexes peuvent être imposées entre des unités musicales de diverses natures.

2. DU SIGNAL À LA SYMBOLIQUE

La composition musicale peut être considérée comme un acte symbolique de projection d'un signal hypothétique

qui n'existe que dans l'esprit d'un compositeur vers une représentation symbolique efficace. Cette approche a été largement utilisée depuis des siècles, car elle permet le partage et la distribution de cette image mentale aux interprètes qui essayeront de le reproduire de la manière la plus exacte. La notation symbolique doit donc être la plus représentative possible afin de maintenir l'idée génératrice aussi proche que possible de l'intention du compositeur. La musique occidentale a ainsi été largement fondée sur ce « paradigme harmonique », c'est à dire qu'elle a construit ses concepts musicaux à partir de sons harmoniques et considérés comme stationnaires. La notation musicale occidentale est également sans coïncidence, cohérente avec cette vision (noter la hauteur d'une note comme un point consiste à réduire un son à sa fondamentale, et considérer qu'à l'intérieur de ce son, les harmoniques suivent une évolution parallèle à celle de la fondamentale), tel que le sont les autres systèmes de musique (polyphonie, tonalité).

Cependant, cette tradition symbolique doit coexister avec une préoccupation des compositeurs de plus en plus marquée pour le timbre et une conception musicale qui consiste à former un substrat et une évolution musicale à partir d'éléments caractérisés, de motifs courts, qui se combinent ou se chevauchent de manière à créer une forme plus holistique. L'évolution des pratiques musicales nous a amenés à reconsidérer les sons harmoniques et bruités comme partie intégrante de notre perception musicale. Ainsi, un des intérêts théoriques majeurs provenant de la composition de mélanges sonores est l'utilisation de représentations multi-échelles pour relier des données hétérogènes. De là, émerge une question contemporaine de la recherche en informatique musicale : l'interaction signal / symbolique. Ces deux courants de recherche sont longtemps restés imperméables l'un à l'autre, en partie en raison de l'apparente hétérogénéité de leurs objets d'étude. D'une part, l'analyse et la synthèse de sons numériques ont contribué à la compréhension du son et ont permis la production de sons inouïs. D'autre part, l'approche algorithmique a mis l'accent sur l'analyse des structures symboliques de la notation musicale. Mais la composition des mélanges sonores se situe précisément à l'intersection de ces deux axes de recherche. Si elle prétend créer des timbres, c'est à nouveau à travers un processus d'écriture. C'est donc un point de rencontre essentiel entre les domaines symboliques et spectraux. Ainsi, cette question ne peut être saisie par la musique assistée par ordinateur seulement à la condition de la convergence de l'approche signal et de l'approche algorithmique.

La question des mélanges sonores pourrait encore être manipulée dans le domaine de l'écriture du timbre instrumental. Mais même si nous pouvons la considérer comme une forme de composition, de par son utilisation extensive du timbre comme élément esthétique central, la portée des mélanges sonores en est beaucoup plus profonde. Le son comme matériau coexiste et interagit avec des structures formelles. De nos jours, les ordinateurs offrent la possibilité de manipuler un objet sonore dans un processus de

composition, tout en étudiant simultanément ses propriétés acoustiques au moyen d'outils analytiques. On peut ainsi relier le domaine symbolique de l'écriture avec le domaine spectral des possibilités du timbre. Il devrait donc être possible pour le compositeur de relier l'exploration du son à l'organisation des données à caractère symbolique.

Cependant, même avec des systèmes de composition modernes, les unités spectrales sont généralement réduites à de simples « blocs » symboliques qui seront placés sur une ligne temporelle. Dans ce schéma, les relations temporelles sur la variabilité du signal et la multidimensionalité du timbre résultant sont complètement perdus, afin de faire face à une facilité ergonomique d'utilisation pour l'interaction homme-machine. Nous présentons ici un nouveau schéma compositionnel dans lequel des relations temporelles peuvent être imposées sur les descripteurs du signal tout autant qu'entre des unités spectrales et symboliques. Comme nous considérons une nouvelle approche de la composition musicale, où les propriétés du timbre sont un élément clé du sens musical, il est nécessaire de modéliser précisément l'évolution temporelle des descripteurs spectraux.

3. MODÉLISATION TEMPORELLE

La musique est un processus articulé, ce qui comprend aujourd'hui non seulement la définition des structures macro-temporelles, mais aussi une compréhension plus fine des propriétés micro-temporelles du timbre. Par conséquent, si nous intégrons le timbre comme élément central du processus de composition, nous devons nécessairement étendre sa compréhension afin de composer à des échelles temporelles supérieures. Le territoire de timbre n'est pas limité à une structure statique des proportions. Il est plutôt composé de « lois de variation » qui régissent l'interaction entre les fonctions de fréquence et d'amplitude dans un contexte évolutif dans le temps. Il est donc indispensable de passer à un niveau supérieur de modélisation, par la compréhension des qualités micro-temporelles du timbre afin de capturer le son comme un structure spectro-temporelle complète. Nous proposons donc une modélisation temporelle aussi bien des descripteurs spectraux que des harmoniques par l'utilisation de Gaussian Mixture Models (GMM) [18]. La spécificité de notre modèle est que les moyennes des fonctions gaussiennes sont espacées par une distance proportionnelle à un paramètre de diffusion. Ce paramètre représente également l'écart-type commun à toutes les distributions, conduisant ainsi à un GMM à variance fixe (GMMFV). Ce modèle permet une grande réduction de la quantité d'informations pour le stockage dans la base de données SQL tout en conservant une bonne précision globale. Ce modèle permet également une inférence automatique des propriétés d'un son identique de longueur modifiée.

3.1. Notre modèle

Lors de l'établissement d'un modèle, il est toujours nécessaire d'effectuer un compromis entre la précision et la réduction de données. Pourtant, face à ce problème d'incertitude, nous pouvons utiliser des études perceptives afin de guider notre discrimination des propriétés inutiles. Une étude particulièrement pertinente dans le contexte des modèles instrumentaux a été menée par McAdams et al. [14] et traite de la discrimination des sons instrumentaux aux paramètres spectro-temporels simplifiés. Cette étude donne un aperçu de l'importance perceptuelle de diverses caractéristiques temporelles et spectrales, et lesquelles peuvent donc être simplifiées. D'abord, deux simplifications semblent extrêmement discriminantes. La cohérence des enveloppes d'amplitude et le lissage des enveloppes spectrales montrent que l'évolution indépendante de l'amplitude des partiels est une caractéristique fondamentale. Ces résultats mettent en évidence le fait que le calcul d'amplitudes moyennes pour chaque partiel est largement insuffisante. Il est donc crucial de représenter l'enveloppe de façon indépendante pour chaque partiel. En revanche, la simplification des microvariations, tant en amplitude qu'en fréquence semble beaucoup moins discriminante perceptivement. Il est donc possible de lisser (jusqu'à un certain niveau) les diverses enveloppes, tout en conservant l'information essentielle.

Pour représenter l'évolution de chaque partiel ou d'un descripteur spectral, nous utilisons les modèles de mélanges gaussiens (GMM). L'évolution d'un partiel est donc définie comme la somme pondérée de distributions gaussiennes

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

avec C le nombre de composants, μ_c la moyenne, σ_c la variance et α_c le poids du composant c , $0 < \alpha_c < 1$ pour tous les composants et $\sum_{c=1}^C \alpha_c = 1$. La liste des paramètres

$$\theta = \{\alpha_1, \mu_1, \sigma_1, \dots, \alpha_C, \mu_C, \sigma_C\} \quad (2)$$

défini donc un modèle particulier de mélange gaussien. Le problème est donc de trouver l'ensemble des paramètres $\hat{\theta}$ qui maximise la probabilité $\mathcal{L}(X; \theta)$ d'observer les données

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} (\ln \mathcal{L}(X; \theta)) \quad (3)$$

Afin d'estimer les paramètres du modèle, nous utilisons l'algorithme Expectation Maximization (EM). La manière dont cet algorithme est appliqué aux mélanges gaussiens [9] est la suivante. Les données connues X sont interprétées comme des données incomplètes. La partie manquante Y consiste à savoir quel composant 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 est produit par le composant c , ou zéro autrement. L'étape

d'expectation consiste à calculer l'espérance conditionnelle de la log-vraisemblance de toutes les données. La probabilité peut être calculée en utilisant la loi de Bayes :

$$w_{n,c} = \frac{\alpha_c^i p(x_n | c; \theta^i)}{\sum_{j=1}^C \alpha_j^i p(x_n | j; \theta^i)} \quad (4)$$

où α_c^i est la probabilité a priori (de l'estimation θ^i) et $w_{n,c}$ est la probabilité a posteriori que x_n soit produit par le composant c . La particularité du modèle GMMFV est que les moyennes des fonctions gaussiennes sont espacées d'une distance proportionnelle au paramètre de diffusion σ_k qui est également l'écart-type commun à toutes les distributions. On obtient donc le mélange suivant :

$$B_{x;\theta} = \sum_{c=1}^C \alpha_c \mathcal{N}(x; \mu_k + c \cdot \sigma_k, \sigma_k) \quad (5)$$

Il suffit donc de conserver uniquement la moyenne μ_k de la première gaussienne et l'écart-type commun σ_k . En appliquant l'étape de maximisation pour estimer la distribution des paramètres pour un mélange de C composants, on obtient les formules de mise à jour

$$\alpha_c^i = \frac{1}{N} \sum_{n=1}^N w_{n,c} \quad (6)$$

$$\mu_k^i = \frac{\sum_{c=1}^C \sum_{n=1}^N (x_n - c \cdot \sigma_k^{i-1}) w_{n,c}}{\sum_{c=1}^C \sum_{n=1}^N w_{n,c}} \quad (7)$$

$$\sigma_k^i = \frac{1}{2} \left(\left((a_k^i)^2 + 4b_k^i \right)^{1/2} - a_k^i \right) \quad (8)$$

$$\begin{cases} a_k^i = \sum_{c=1}^C \sum_{n=1}^N c (x_n - \mu_k^i) w_{n,c} \\ b_k^i = \sum_{c=1}^C \sum_{n=1}^N (x_n - \mu_k^i)^2 w_{n,c} \end{cases} \quad (9)$$

Les nouvelles estimations sont réunies dans θ^i . Les équations sont évaluées à nouveau avec ces nouvelles estimations jusqu'à ce que le critère de convergence soit satisfait.

3.2. Modification de durée

La plupart des instruments de l'orchestre produisant des sons entretenus, il semble obligatoire qu'une modification de durée puisse être faite sur les descripteurs des sons instrumentaux. En utilisant les paramètres du modèle, il devient automatique d'en déduire la structure d'un son instrumental identique, mais de durée différente. Nous considérons la modification par expansion ou compression de l'enveloppe temporelle sur la partie entretenue des sons uniquement. Ainsi les segments d'attaque et de relâchement sont conservés identiques par le calcul de leurs emplacements respectifs et en appliquant ensuite les changements uniquement aux paramètres appartenant à la partie entretenue. Ce choix provient du fait que l'excitation durant l'attaque et le relâchement est généralement identique quelle que soit la durée de la note. Les gaussiennes modifiées sont donc celles dont la moyenne satisfait

$$pos_{att} + \epsilon < \mu_k < pos_{rel} - \epsilon \quad (10)$$

avec pos_{att} la fin du segment d'attaque, pos_{rel} le début du segment de relâchement et ϵ un facteur de tolérance temporelle. On obtient ainsi l'ensemble de gaussiennes à modifier :

$$k \in K = \{\mathcal{N}(\mu_k, \sigma) / pos_{att} + \epsilon < \mu_k < pos_{rel} - \epsilon\} \quad (11)$$

Deux méthodes différentes peuvent alors être appliquées afin de modifier avec précision la durée de la structure.

3.2.1. Modification par dilatation

Dans notre modèle, toutes les gaussiennes sont reliées entre elles par un paramètre de diffusion commun σ_k . Ce paramètre peut être utilisé d'une manière simple pour réaliser l'expansion ou la compression sur le mélange de gaussiennes. Si nous voulons changer la longueur du son par un facteur γ , il nous suffit de multiplier la variance de chaque gaussienne par ce facteur, soit $\sigma'_k = \gamma\sigma_k$. Nous devons ensuite mettre à jour les poids des gaussiennes en conséquence afin de compenser la modification de variance $\alpha'_k = 2 \cdot \gamma \cdot \alpha_k$ puis normaliser tous les poids. Les moyennes des gaussiennes étant reliées par un facteur de diffusion commun, il suffit de calculer les moyennes pour $n \geq \min(k)$

$$\mu_n = \mu_{n-1} + (\sigma'_{n-1} + \sigma'_n) \quad (12)$$

3.2.2. Modification par répétition

Il est possible pour des cas particuliers comme le vibrato et le trémolo d'effectuer une modification plus pertinente par répétition. Cela est fait en conservant les paramètres d'origine mais en changeant le nombre de gaussiennes, puis en mettant à jour les poids des composants correspondants. Nous devons d'abord calculer le nouveau nombre de gaussiennes

$$N_{new} = \lceil \gamma \cdot N_{k \in K} \rceil + N_{g \notin K} \quad (13)$$

Les paramètres des composants pré-existants restent inchangés, et nous considérons que le poids des composants nouvellement ajoutés suit un comportement cyclique

$$\alpha_n = \alpha_{n - \text{card}(K)} \quad (14)$$

Il est nécessaire de noter que les deux approches sont applicables à des cas spécifiques. En effet, s'il s'agit d'une enveloppe linéairement décroissante, la modification par répétition apparaît irréaliste. Inversement, pour les trémolos ou vibratos, même si la modification par dilatation semble suggérer une structure valide, il est évident que l'approche par répétition est beaucoup plus pertinente. Ainsi, il serait optimal de choisir la méthode de modification de durée en suivant une taxonomie des échantillons en cours d'analyse.

3.3. Résultats

Nous discutons ici des résultats de notre modèle. Pour cela, nous le comparons avec les polynômes de Legendre et les GMM à variance libre. Il est possible d'observer leur performance sous des angles différents. Nous soulignons l'importance de la précision du modèle en calculant l'écart entre le spectre d'origine et un spectre resynthétisé à partir de données réduites. Cette précision est confrontée à la réduction de données induites par le modèle. Cela nous permet d'évaluer la performance du modèle à saisir les caractéristiques essentielles d'un son.

3.3.1. Précision

La précision est comparée en utilisant deux mesures d'erreur différentes, la norme L^2 et la divergence de *Kullback-Leibler* généralisée. On calcule celles-ci entre le spectre d'origine et le spectre resynthétisé. Toutefois, la durée et le contenu des échantillons pouvant être très variables, on normalise ces mesures avec la divergence entre chaque échantillon et une période équivalente du silence. De cette façon nous pouvons comparer les résultats avec des sons de natures diverses. Enfin nous ajoutons à ces résultats le temps de resynthèse ainsi que la réduction de données induites par les modèles. Le tableau 1 résume les résultats de l'analyse sur les 21.381 échantillons de la base de données.

	Polynômes de Legendre	Gaussian Mixture Models	
		Variance libre	Variance fixe
Synthèse	0.003 +/- 0.004	0.021 +/- 0.003	0.025 +/- 0.013
L^2	39.7 +/- 836.9	19.43 +/- 218.3	8.307 +/- 37.34
$L^2/Sil.$	0.57 +/- 0.152	0.214 +/- 0.027	0.089 +/- 0.022
KL	335.7 +/- 6987	272.8 +/- 458.9	201.7 +/- 362.7
$KL/Sil.$	0.97 +/- 0.504	0.618 +/- 0.004	0.381 +/- 0.015
Réduction	77.2 +/- 498.7	52.20 +/- 339.3	48.68 +/- 246.6

Table 1. Valeur moyenne et variance pour le temps de resynthèse, la réduction de donnée et la précision des différents modèles.

Comme nous l'espérions, nous obtenons les meilleurs résultats avec le modèle à variance fixe qui permet d'obtenir la meilleure précision et ce pour une réduction de données équivalente. La précision du modèle à variance libre pourrait être accrue en augmentant le nombre de composants mais cela augmenterait proportionnellement les chances d'obtenir des estimations singulières.

4. BASE DE DONNÉES SPECTRALE

Durant les dernières décennies, un grand nombre de bases de données ont été établies afin de faire face à l'incroyable diversité et la variabilité des possibilités instrumentales. Les bases de données les plus importantes étant *Studio On Line* (SOL) [4], *Real World Computing* (RWC) [10], et *Vienna Symphonic Library* (www.vsl.co.at).

Cependant, celles-ci manquent clairement d'une organisation architecturale car elles sont uniquement composées de fichiers sonores et de répertoires sans schéma organisationnel complexe. Nous présentons ici une première base de données structurée qui intègre des informations symboliques, spectrales et textuelles sur les fichiers sonores contenus dans toutes les bases précédemment citées. Notre base de données repose sur une architecture SQL et contient pour chaque son des informations sur les descripteurs symboliques et spectraux calculés avec *IRCAM-Descriptor* [16] puis modélisés temporellement grâce au modèle GMMFV.

Cette base de données permet d'effectuer des recherches rapides basées aussi bien sur des descripteurs spectraux que sur de l'information symbolique simultanément. En conséquence, le système est capable de proposer des solutions basées sur des critères à la fois spectraux et symboliques. De plus le système de requêtes temporelles permet de trouver intuitivement des sons correspondant à des courbes d'évolution des descripteurs tels que l'amplitude ou la fréquence fondamentale.

4.1. Architecture SQL

Afin d'organiser et d'accélérer les requêtes sur une large base de données sonores, nous avons conçu une architecture SQL simple et efficace. Comme nous l'avons mentionné précédemment, chaque son de la base de données est représenté par des informations textuelles, symboliques ainsi que des descripteurs spectraux. Pour cela, chaque son nécessite une description qui peut contenir pour les instruments le mode de jeu, hauteur, intensité et autres descriptions informatives. Nous analysons ensuite chaque son en utilisant l'ensemble des descripteurs disponibles dans *IRCAMDescriptor* afin d'obtenir autant d'informations spectrales que possible. Chaque descripteur est ensuite modélisé temporellement en utilisant le modèle présenté dans la section précédente. Les descripteurs représentant la moyenne, l'écart-type et la modélisation temporelle sont stockés dans la base de données. Cette approche permet de faire des recherches conjointes sur des informations issues de la symbolique ou du signal.

Les informations contenues dans la base de données peuvent être de natures diverses :

- Les valeurs atomiques qui représentent une seule information. Les informations symboliques sont stockés avec des chaînes de caractères de longueur variable. La majorité des descripteurs moyens et d'écart-type étant stockés sous forme de valeurs à virgule flottante.
- Les tableaux sont utilisés pour certains des descripteurs moyens qui sont constitués de valeurs multiples (par exemple fréquence moyenne des partiels, platitude spectrale par sous-bande)
- Les modélisations GMMFV sont stockées avec les informations sur la moyenne de la première gaussienne, le facteur de diffusion commun et une liste des amplitudes.

- Les tableaux de gaussiennes sont utilisés pour stocker la modélisation temporelle de descripteurs dont la valeur moyenne est représentée par un tableau.

Pour chaque type d'information, un ensemble spécifique de relations est disponible. Il est donc possible de faire des requêtes simultanément sur des ensemble de données hétérogènes.

4.2. Requêtes temporelles

Pour permettre aux utilisateurs d'effectuer des recherches intuitives dans le lexique, nous introduisons un système de requêtes temporelles. Celui-ci peut être couplé avec n'importe quel type de requête symbolique ou statique et permet ainsi de trouver rapidement et efficacement des sons correspondant à des contraintes complexes. Il permet ainsi de définir des critères d'évolution temporelle sur tout sous-ensemble des descripteurs disponibles. L'utilisateur peut ainsi chercher des sons correspondant à des idées complexes d'évolution. Le principe de requêtes temporelles sur la base de données peut être placé dans le contexte plus global de la *similarité des séries temporelles*. Ce problème peut être ainsi découpé en trois composantes principales :

- La *représentation* des données qui va déterminer la manière dont la recherche et le stockage seront effectués.
- La mesure de *similarité* qui permettra de définir une hiérarchie de ressemblance entre les séries temporelles.
- La technique d'*indexation* des séries qui permet d'organiser les données pour accélérer la recherche dans de grandes bases.

Dans un premier temps, la représentation des données doit fournir un type compact et permettant une recherche rapide. Cette représentation doit également permettre d'évaluer une relation de similarité entre les séries. Un grand nombre de représentations pour la similarité de séries temporelles ont été introduites, notamment la transformée de Fourier discrète [8], la transformée en ondelettes [7], les modèles d'approximation par morceaux [15] et la décomposition en valeurs singulières [11]. Des travaux récents suggèrent que ces modèles ne présentent que peu de différences en termes de possibilités d'indexation [12]. Cependant de nombreuses différences existent quant à la qualité des résultats ainsi que la rapidité des requêtes effectuées sur la base de données. Une propriété principale de toutes les représentations précédentes est qu'elle sont à valeurs réelles. Celle-ci limite donc les algorithmes, structures de données et définitions disponibles pour ces représentations. Ces limitations ont amené à considérer l'utilisation de représentations symboliques pour les séries temporelles. Nous utilisons donc comme représentation un dérivé du système SAX (Symbolic Aggregate approximation) [13]. Celui-ci permet de réduire une série temporelle de longueur arbitraire n à une chaîne de caractères de longueur w (typiquement $w \ll n$) suivant un alphabet de taille a . Cette représentation permet donc d'obtenir une version symbolique de taille très réduite des différentes sé-

ries temporelles. La série à représenter est découpée suivant des points uniformément espacés puis chaque unité est transformée en un caractère en fonction de la valeur moyenne de la série à cet instant. Notre décomposition prends également en compte la valeur moyenne de la dérivée pour chaque unité.

Cette représentation symbolique permet de définir une fonction de similarité sur des chaînes de caractères, prenant en compte la distance entre les différents symboles. Ainsi la similarité entre requêtes temporelles peut dépasser le schéma classique de la distance euclidienne qui ne prend pas en compte des critères perceptuels. En effet, les humains sont capables de détecter des similarités à différentes échelles sur une courbe. En observant de manière simultanée les caractéristiques d'une série, nous sommes capables de nous abstraire des problèmes d'amplitude, de mise à l'échelle temporelle et de bruit. Il est évident qu'une distance point-à-point telle que la distance euclidienne est incapable d'un tel niveau d'abstraction. Une mesure de similarité entre séries temporelles devrait ainsi être invariante suivant une famille de transformations. Soit une série temporelle représentée par la fonction $f(t)$, nous considérons les transformations suivantes :

- **Décalage** : La série $g(t)$ obtenue par décalage linéaire de la série originelle $g(t) = f(t) + k$ avec k une constante.
- **Amplification uniforme** : La série $g(t)$ obtenue par multiplication de la série $g(t) = k.f(t)$ avec k une constante.
- **Modification temporelle uniforme** : La série $g(t)$ obtenue par modification constante de l'échelle temporelle $g(t) = f(kt)$ avec k une constante.
- **Amplification dynamique** : La série $g(t)$ obtenue par multiplication de la série par une fonction d'amplitude $g(t) = h(t).f(t)$ avec $h(t)$ une fonction telle que pour chaque t , $h'(t) = 0$ si et seulement si $f'(t) = 0$.
- **Modification temporelle dynamique** : La série $g(t)$ obtenue par modification dynamique de l'échelle temporelle $g(t) = f(h(t))$ avec $h(t)$ une fonction positive et strictement croissante.
- **Bruit additif** : La série $g(t)$ obtenue par ajout d'une composante bruitée $g(t) = f(t) + \epsilon(t)$ avec $\epsilon(t)$ un bruit blanc indépendant identiquement distribué.

Ces transformations pouvant être combinées simultanément pour former de nouvelles transformations. Pour parler au problème de décalage et d'amplification uniforme, toutes les séries sont d'abord décalées et amplifiées de manière à ce que leur moyenne soit nulle et que leur déviation standard soit unitaire. Le bruit additif est alors implicitement géré par le passage à la représentation symbolique. Enfin nous utilisons une mesure de similarité [13] sur les chaînes de caractères résultantes permettant de prendre en compte les transformations présentées précédemment. Soit deux chaînes Q et C de longueur w , la chaîne C représentant une série de taille originelle n , la similarité

entre celles-ci est définie par

$$D(Q, C) = \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^w (dist(q_i, c_i))^2}$$

avec q_i et c_i le i^{eme} caractère des chaînes correspondantes et $dist(a, b)$ une fonction permettant d'obtenir la distance entre les symboles a et b . Grâce à cette mesure, il est alors possible d'obtenir la similarité entre une série donnée en requête et les séries présentes dans la base de données. L'utilisateur peut alors chercher intuitivement des sons dans la base dont les descripteurs suivent une évolution particulière. Ainsi il suffit de dessiner une courbe pour trouver les sons dont les descripteurs correspondent à celle-ci. Il est alors facilement possible de transposer une idée sonore complexe en recherche intuitive.

4.3. Serveur OSC

Pour que cette base de données puisse être utilisable dans tout type de situations, nous avons développé un serveur OSC qui peut être utilisé comme une interface entre un client et la base de données SQL. Ce serveur a été développé en Matlab et permet à plusieurs clients de se connecter simultanément. Le serveur peut être utilisé pour se connecter à partir d'un client distant sur une base de données distante ou sur une architecture locale. Un ensemble de fonctions encapsule les principales fonctionnalités de la base de données. De cette façon, les clients peuvent ajouter des sons ou effectuer des requêtes sur des descripteurs spécifiques d'une manière totalement transparente. Pour des descripteurs complexes telle que la modélisation GMMFV, le serveur récupère les valeurs du modèle et synthétise la fonction temporelle complète. De cette façon, le client ne voit que la forme définitive des descripteurs. Il est également possible pour le client de demander un aperçu du son. En utilisant la modélisation appliquée à l'évolution des partiels, le serveur est en mesure de calculer une version approximative de l'onde sonore correspondante. Il est enfin possible de sérialiser la base de données entière dans un seul fichier afin de l'utiliser hors connexion.

5. MAQUETTES SPECTRALES

Nous allons voir dans cette section, l'interface pour la base de données spectrale décrite à la section 4. Celle-ci est implémentée comme une bibliothèque dans le logiciel OpenMusic. Tout d'abord nous avons défini une nouvelle classe de sons appartenant à la base de données (`db-sound`). L'utilisateur peut faire des instances de cette classe en envoyant des requêtes SQL à la base de données spectrale (voir Figure 1). Plus qu'un son, un `db-sound` est une collection de sons dont la cardinalité est spécifiée par le paramètre `max-solutions` de la requête. Comme les fichiers sonores classiques, les instances `db-sound` sont des citoyens de première classe (c'est à dire qu'ils peuvent être inclus dans les programmes visuels, notes ou

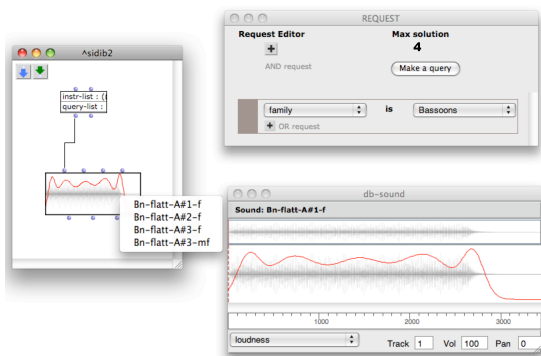


Figure 1. Création d'une instance de la classe `db-sound` par une requête SQL. Dans cet exemple, nous cherchons des sons appartenant à la famille des bassons. La nouvelle instance `db-sound` contient quatre sons. Un éditeur spécial permet d'afficher différents descripteurs de ceux-ci.

d'autres éditeurs OpenMusic). En utilisant des instances `db-sound` ainsi que les outils généraux de programmation visuelle de l'environnement OpenMusic, les compositeurs sont ainsi pourvus de moyens pour développer des processus complets en ce qui concerne la question des mélanges sonores. Dans ce contexte, les structures du programme fonctionnel les laissent gérer la complexité en maintenant un contrôle hiérarchique d'abstractions musicalement pertinentes jusqu'aux processus de synthèse.

En particulier, nous avons mis en place une nouvelle classe de maquettes appelées *Spectral Maquettes*. La maquette est une extension de la notion de programme visuel avec ajout d'autres dimensions spatiales et temporelles, permettant de mettre les éléments du cadre de composition (structures de données et processus) en relation étroite avec ces deux dimensions. Dans un éditeur de maquette, les cases (appelées boîtes temporelles) représentent des unités fonctionnelles (programmes) qui produisent des réalisations musicales (voir Figure 2). La position et les propriétés graphiques de ces boîtes sont associées à un sens temporel et structurel, surtout, l'axe horizontal de l'éditeur représente le temps, afin que la position et l'extension horizontale puissent être liées à des terminaisons et des durées.

Les boîtes temporelles peuvent également être liées par des liaisons fonctionnelles, afin que la maquette entière puisse finalement être considérée comme un programme, comprenant une sémantique fonctionnelle et temporelle. Les relations temporelles et les contraintes (synchronisation, terminaisons relatives, etc) peuvent donc être imposées entre les unités en définissant les paramètres temporels dans leurs programmes correspondants. Par conséquent, le calcul permet de déterminer la structure temporelle. En outre, la possibilité d'intégrer des maquettes dans une maquette permet de construire des structures hiérarchiques temporelles.

La maquette spectrale est une extension de la maquette classique visant à fonder la relation du matériau musical et des processus avec les propriétés micro-temporelles du timbre. En effet, les mélanges sonores exigent d'ac-

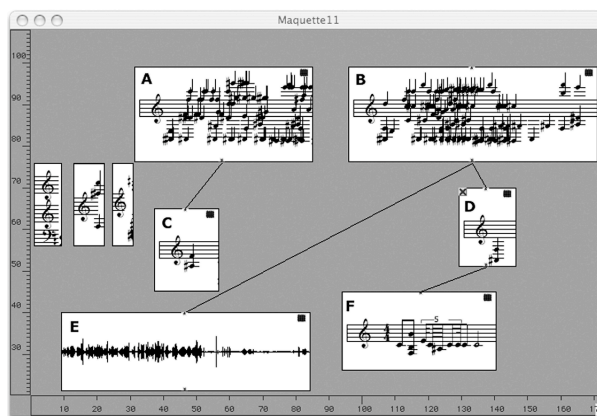


Figure 2. Un éditeur de maquette. Chaque unité représente un programme qui permet le calcul d'objets musicaux, intégré dans une référence temporelle suivant l'axe horizontal. (L'axe vertical est constitué d'unités arbitraires et sans dimension pour permettre une meilleure organisation spatiale.) Les connexions fonctionnelles sont visibles entre les entrées / sorties de certains de ces programmes.

order une attention particulière à l'évolution temporelle des composantes individuelles ainsi que l'organisation, le séquençage, et les articulations d'une forme musicale globale. Nous montrons dans la Figure 3 un exemple de maquette spectrale. Quand un `db-sound` est mis dans la maquette, la boîte sous-jacente montre la forme d'onde du son augmentée avec plusieurs informations comme des points clef (par exemple *attaque*, *sustain* et *release*) ou l'évolution de certains descripteurs (par exemple, *loudness*, *noisiness*, centroïde spectral, etc.). Les unités à l'intérieur d'une maquette spectrale peuvent être reliées entre elles de différentes manières. Dans la Figure 3 (a) nous pouvons voir comment le démarrage d'un son est lié à la première note d'une mélodie; déplacer une de ces unités force le déplacement de l'autre, afin de respecter cette contrainte. Dans la partie (b) trois sons se superposent, on peut voir dans cet exemple plusieurs types de relations temporelles : le plus bas est attaché à un marqueur dans la *timeline* globale, ce son démarrera toujours à partir de ce moment, la boîte du milieu quant à elle, se termine en même temps que la plus basse (d'autres relations de Allen peuvent aisément être imposées entre les boîtes), et enfin les relations entre la boîte la plus haute et celle du milieu sont définies au moyen de marqueurs internes aux sons (ces points peuvent être définis à la main ou provenir de requêtes sur les bases de données). Dans la partie (d) le `db-sound` est considéré comme l'entrée d'un programme qui génère une séquence de notes dont le profil mélodique suit l'évolution de l'amplitude du son. Enfin, dans la partie (e) un visualisateur permet de sélectionner un descripteur particulier et calculer le résultat de la superposition de tous les `db-sounds` dans la maquette pour celui-ci. Dans un premier temps ce visualisateur permet de voir l'évolution globale de la superposition, mais il n'est pas difficile d'imaginer qu'en changeant la courbe, l'uti-

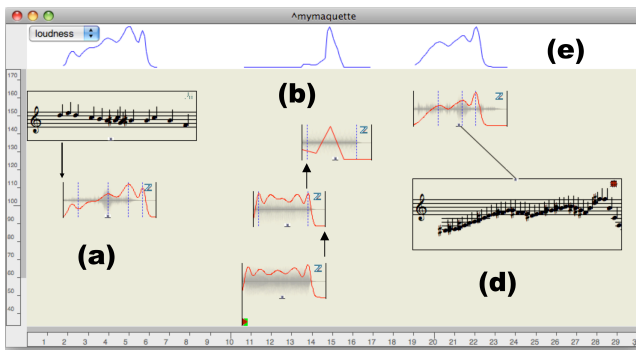


Figure 3. Exemple concret de l'utilisation des maquettes spectrales. De multiples relations fonctionnelles, macro et micro-temporelles sont définies entre les unités.

lisateur puisse lancer une requête qui recherche de nouvelles unités permettant d'approcher cette nouvelle courbe. Dans l'exemple précédent, nous avons essayé de montrer comment les maquettes spectrales permettent de structurer les informations musicales à différents niveaux :

- Le niveau statique de la forme, à savoir l'organisation des unités dans le temps.
- Le niveau dynamique et paradigmatique de la forme (c'est à dire les contraintes temporelles entre les unités).
- Le niveau syntaxique, c'est à dire le calcul permettant la construction du discours musical entre les unités temporelles.
- Le niveau du matériau, c'est à dire les db-sounds qui participent à la maquette.

Ces quatre niveaux d'information sont évidemment liés. L'avantage le plus important dans le concept de maquette spectrale est d'offrir une visualisation de cette interaction et dans le même temps, un contrôle interactif de celle-ci. Cela produit une source d'exploration et d'expérimentation :

- La recombinaison au niveau de la forme globale. Les unités temporelles sont déplacées et étendues dans le temps sans changer les trois autres niveaux.
- Modification des relations fonctionnelles. Nous ne modifions pas la position des blocs, mais leur relation de causalité.
- Modification de la syntaxe. Les algorithmes construisant le matériau sonore peuvent être changés en fonction des objectifs de composition.
- Modification du matériau en lançant de nouvelles requêtes à la base de données.

En combinant ces différentes procédures, des expérimentations musicales sophistiquées peuvent être effectuées. Il est important de noter que dans les maquettes spectrales, l'utilisateur ne travaille pas directement avec des sons. Le son résultant d'une maquette spectrale est synthétisé par le serveur afin de créer un nouveau son de la base ou simplement d'être écouté.

6. CONCLUSION

Nous avons présenté dans cet article un nouveau système permettant d'explorer l'univers combinatoire des mélanges sonores à des fins compositionnelles. En s'appuyant sur une architecture SQL, il est possible de faire des requêtes rapides basées simultanément sur des critères symboliques et spectraux. Un système de requêtes basé sur la forme temporelle de descripteurs a été mis en œuvre afin d'ajouter la possibilité de rechercher des évolutions de descripteurs spécifiques à l'intérieur de la base. Cette approche permet de trouver des unités qui correspondent à des idées sonores complexes. La modélisation temporelle appliquée à chaque descripteur spectral ainsi qu'aux différents partiels permet de préserver l'ensemble de la structure spectro-temporelle. Ce lexique organisé a été utilisé comme une base pour le système de maquettes spectrales. Les maquettes permettent des interactions temporelles et structurelles complexes qui peuvent être imposées entre des unités musicales de diverses natures. Ce système devrait permettre des possibilités d'exploration en profondeur du matériau sonore grâce à l'accès à la multidimensionalité de timbre et à la complexité des mélanges sonores.

Plusieurs directions de recherche peuvent être entreprises sur les maquettes spectrales. Premièrement, s'agissant du modèle temporel, le type de distributions peut être modifié afin d'augmenter la précision globale et la réduction des données du modèle. Le modèle de mélange gaussien asymétrique [2] pourrait être appliqué et renforcerait la réduction des données. Pour l'architecture SQL, un système d'indexation efficace permettrait d'imaginer la mise à l'échelle de la base de données à l'intégralité des collections sonores disponibles. Enfin, l'intégration du logiciel d'orchestration *Orchidée* [6] développé à l'Ircam permettrait des possibilités d'exploration supplémentaires. Ce système est basé sur l'idée de son *cible* qui doit être reproduit en faisant une recherche de combinaisons sonores au sein de grandes banques d'échantillons instrumentaux. La recherche est effectuée par un algorithme génétique qui permet une recherche multi-objectif vers un timbre cible, dans laquelle plusieurs dimensions perceptives sont optimisées conjointement. Il deviendrait ainsi possible pour le compositeur de dessiner des formes de descripteurs et le système ne proposerait pas seulement une solution unique mais plusieurs unités résultant en un mélange sonore correspondant aux différents critères.

7. REFERENCES

- [1] C. Agon, G. Assayag, O. Delerue, and C. Rueda, "Objects, time and constraints in OpenMusic," *Proceedings of the International Computer Music Conference*, Ann Arbor, USA, 1998, pp. 406 – 415.
- [2] R. B. Arellano-Valle, H. Bolfarine, and V. H. Lachos, "Skew-normal linear mixed models," *Journal of Data Science*, vol. 3, pp. 415 – 438, 2005.
- [3] G. Assayag, C. Rueda, M. Laurson, C. Agon, and O. Delerue, "Computer-assisted composition at Ir-

- cam: From Patchwork to OpenMusic,” *Computer Music Journal*, vol. 23, no. 3, pp. 59 – 72, 1999.
- [4] G. Ballet, R. Borghesi, P. Hoffmann, and F. Levy, “Studio online 3.0 : An internet “killer application” for remote access to ircam sounds and processing tools,” in *Actes des Journées Informatique Musicale*, Paris, France, 1999.
- [5] P. Boulez, “Le timbre et l’écriture, le timbre et le langage,” in *Le timbre, métaphore pour la composition*. J.B. Barrière, 1991, pp. 541 – 549.
- [6] G. Carpentier, “Approche computationnelle en orchestration musicale - optimisation multicritère sous contraintes de combinaisons instrumentales dans de grandes banques de sons,” Ph.D. dissertation, University UPMC Paris 6, 2008.
- [7] K. Chan and A. Fu, “Efficient time series matching by wavelets,” in *Proceedings of the 15th IEEE International conference on data engineering*, Sydney, Australia, 1999, pp. 126 – 133.
- [8] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, “Fast subsequence matching in time-series databases,” *SIGMOD Record*, vol. 23, pp. 419 – 429, 1994.
- [9] M. A. T. Figueiredo and A. K. Jain, “Unsupervised learning of finite mixture models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381 – 396, 2002.
- [10] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, “Rwc music database : music genre database and musical instrument sound database,” in *Proceedings of the 4th International Conference on Music Information Retrieval*, Washington, USA, 2003, pp. 229 – 230.
- [11] E. Keogh, K. Chakrabarti, and M. Pazzani, “Locally adaptive dimensionality reduction for indexing large time series databases,” in *Proceedings of ACM SIGMOD conference on management of data*, Santa Barbara, USA, 2001, pp. 151 – 162.
- [12] E. Keogh and S. Kasetty, “On the need for time series data mining benchmarks : a survey and empirical demonstration,” in *Proceedings of the 8th ACM SIGKDD International conference on knowledge discovery and data mining*, Edmonton, Alberta, Canada, 2002, pp. 102 – 111.
- [13] J. Lin, E. Keogh, L. Wei, and S. Lonardi, “Experiencing SAX: a novel symbolic representation of time series,” *Data Mining and Knowledge Discovery*, vol. 15, no. 2, pp. 107–144, 2007.
- [14] S. McAdams, J. W. Beauchamp, and S. Meneguzzi, “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.
- [15] G. P., “Pattern extraction for time series classification,” in *Proceedings of the 5th European conference on principles of data mining and knowledge discovery*, Freiburg, Germany, 2001, pp. 115 – 127.
- [16] G. Peeters, “A large set of audio features for sound description (similarity and classification) in the cuidado project,” IRCAM, Paris, Tech. Rep., 2004.
- [17] P. Schaeffer, *Traité des objets musicaux*, Seuil, 1966.
- [18] J. J. Verbeek, N. Vlassis, and B. Krose, “Efficient greedy learning of gaussian mixture model,” *Neural Computation*, vol. 15, no. 2, pp. 469 – 485, 2003.