

Télécom ParisTech
LAM
Master ATIAM
Rapport de stage

Maître de stage : Gaël Richard et Laurent Daudet
Tuteur universitaire : Emmanuel Saint-James

Enrichissement spectral en audio

Benoit Fuentes

Paris, le 22 juin 2009

Table des matières

1	Introduction	5
2	Etat de l'art	7
2.1	Principe général	7
2.2	Création d'énergie en bande haute	8
2.2.1	Filtrage non-linéaire	8
2.2.2	Techniques de réplication spectrale	9
2.3	Modelage de l'énergie	10
2.3.1	Modèles simples d'enveloppe	10
2.3.2	Modèles statistiques	10
2.3.3	Approche par dictionnaire et quantification vectorielle	10
3	Enrichissement par décomposition atomique	12
3.1	Principe de base	12
3.2	Factorisation en matrices non-négatives (NMF)	13
3.3	Apprentissage du dictionnaire	15
3.3.1	Apprentissage non supervisé	16
3.3.2	Apprentissage supervisé	16
3.4	Enrichissement	17
3.5	Discussion	19
4	Enrichissement du bruit par modèle de signal	20
4.1	Modèle de signal	20
4.2	Algorithme d'enrichissement du bruit	21
4.2.1	Extraction des enveloppes	21
4.2.2	Création du dictionnaire	23
4.2.3	Enrichissement	24
5	Tests et résultats préliminaires	27
5.1	Enrichissement par décomposition atomique	27
5.1.1	Mise en place du test	27
5.1.2	Résultats et discussions	29
5.2	Enrichissement du bruit par modèle de signal	30
5.2.1	Mise en place du test	31

5.2.2	Résultats et discussion	31
6	Conclusion	33
6.1	Résumé du rapport et des contributions apportées	33

Chapitre 1

Introduction

Lors de sa numérisation, de son stockage ou de encore de sa transmission, un signal audio (tel qu'il a été émis par un instrument de musique par exemple) subi généralement de nombreuses perturbations. Par exemple, pour un gain de place ou de rapidité de transmission, il est de plus en plus fréquent de compresser un signal audio (le format de compression le plus populaire étant le format MP3). Dans d'autres domaines d'application comme la téléphonie, on réduira volontairement la bande de fréquence du signal de 0 à 4000 Hz. Dans tout les cas, ces transformations peuvent être assimilées à de la perte d'information, et donc de qualité audio.

Parmi ces pertes de qualité, on trouve souvent une disparition des fréquences les plus aiguës. Or, il est possible dans une certaine mesure de recréer de l'information dans les hautes fréquences à partir des seules fréquences graves. C'est le sujet de ce stage. Pour des raisons de simplicité, nous nous restreignons à enrichir des signaux dits bande basse, qui ne possèdent pas d'énergie pour les fréquences supérieures à 4000 Hz, afin qu'ils deviennent des signaux dits pleine bande, possédant de l'énergie sur toute la bande de fréquence (0-8000Hz). Nous travaillons donc avec des signaux échantillonnés à 16000 Hz.

Après avoir dressé un état de l'art des techniques d'enrichissement spectral (chapitre 2), nous étudierons deux nouvelles approches. La première, décrite au chapitre 3, s'appuie sur une technique de décomposition atomique du signal : le spectre de puissance d'un signal audio de quelques millisecondes peut être décomposé comme une somme pondérée de spectres de base, appelés atomes. La première étape est de créer un dictionnaire d'atomes qu'il faut apprendre sur une base de données de signaux de bonne qualité (c'est-à-dire un signal pleine bande), nous dirons que ce sont des atomes pleine bande. Ces atomes pourront alors être utilisés pour étendre la bande d'un signal bande basse. Le chapitre 4 décrit quant à lui une méthode fondée sur la séparation du bruit des sinusoïdes. Les parties aléatoires (le bruit) et déterministes (les sinusoïdes) sont enrichies séparément et ce chapitre s'intéresse

à l'enrichissement du bruit uniquement. Les tests et les résultats sont discutés dans le chapitre 5 et nous concluons et résumerons les contributions fournies pendant ce stage dans le chapitre 6.

Chapitre 2

Etat de l'art

2.1 Principe général

Plusieurs solutions pour l'enrichissement spectral ont déjà été proposées, et même utilisées dans l'industrie. On trouve dans la plupart des articles de la littérature qui traitent de ce problème des méthodes en deux temps. Un premier temps où l'on crée de l'énergie dans les fréquences aiguës, et un deuxième temps où cette énergie est modélée selon une enveloppe spectrale particulière. La figure 2.1 illustre ces étapes. Lors de l'étape de création

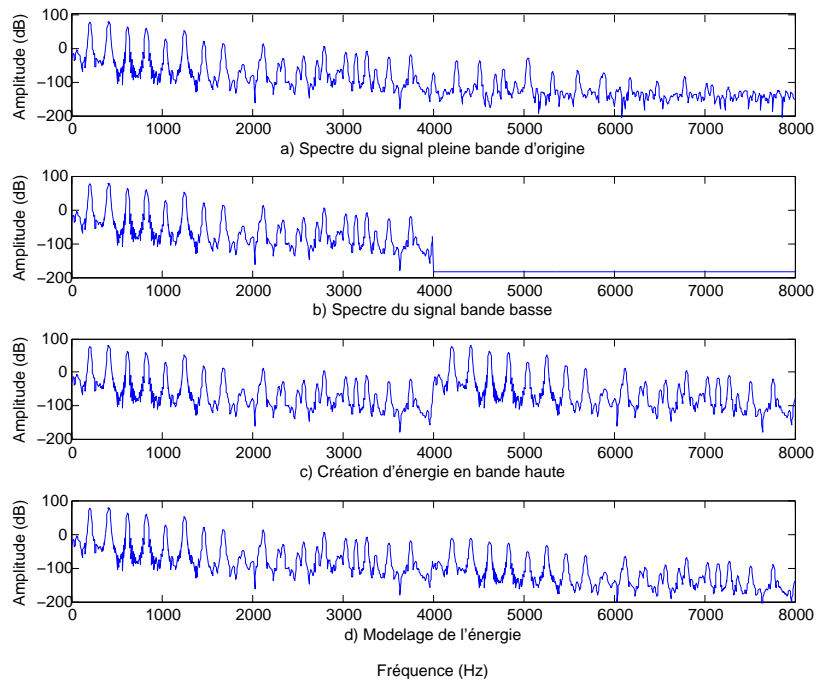


FIGURE 2.1 – Principe général de l'enrichissement spectral

d'énergie, il faut prendre soin de respecter les caractéristiques du signal bande basse, par exemple le caractère harmonique du signal. L'étape de modelage permet quant à elle de respecter l'allure de l'enveloppe spectrale du signal audio. Nous présentons ici un état de l'art des techniques développées pour résoudre chacun de ces deux problèmes.

2.2 Création d'énergie en bande haute

Dans cette section nous nous intéressons aux diverses techniques qui existent pour créer de l'énergie en hautes fréquences en fonction du signal basse bande à enrichir.

2.2.1 Filtrage non-linéaire

Une première manière de créer de l'énergie dans les hautes fréquence est d'appliquer un filtrage non-linéaire sur une partie du signal bande basse. Ce principe est utilisé par exemple dans [7]. Un des effets d'un filtre non linéaire est de générer des harmoniques aiguës multiples des harmoniques graves présentes dans le signal (que l'on suppose harmonique). Un diagramme en bloc sur la méthode complète d'enrichissement de [7] est présenté dans la figure 2.2. Dans ce diagramme, le signal d'entrée est un signal ne comprenant que

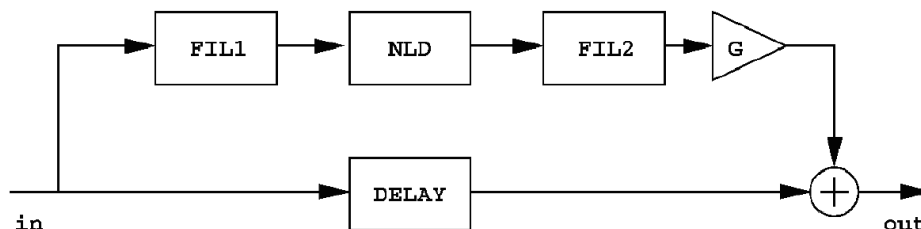


FIGURE 2.2 – Une méthode pour l'enrichissement spectral [7]

des fréquences graves. FIL1 représente un filtre passe bande qui sélectionne les fréquences les plus grave dans le signal d'entrée. NLD (Non Linear Device) est le filtre non linéaire (par exemple un simple filtre redresseur) qui permet la génération de fréquences aiguës. FIL2 est un filtre passe haut qui ne garde que les fréquences les plus aiguës. Enfin, le signal bande haute est multiplié par un gain G puis additionné au signal d'entrée (retardé pour prendre en compte le temps de calcul du traitement du signal haute fréquence).

Ce genre de technique est très simple à développer et très peu coûteux en temps de calcul, mais peut éventuellement introduire des harmoniques non souhaitées dans le signal.

2.2.2 Techniques de réplique spectrale

Afin de créer de l'énergie dans les hautes fréquences, des techniques dites de réplique spectrale ont également été développées. Le principe est simplement de recopier le spectre des basses fréquences dans la zone des hautes fréquences, soit par translation [2], soit par symétrie [1]. La figure 2.3 illustre la réplique spectrale.

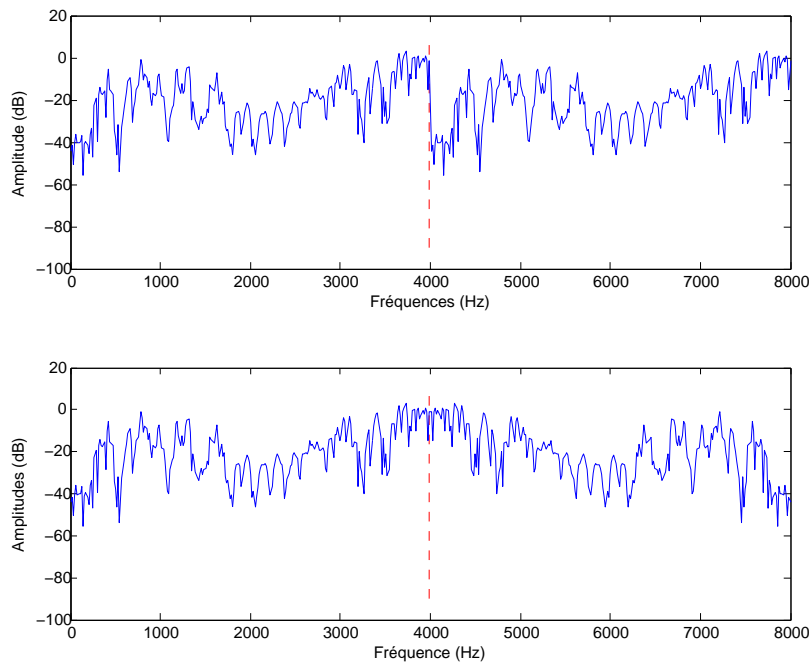


FIGURE 2.3 – Spectre d'amplitude d'après deux méthodes différentes de réplique spectrale

Le coefficient de translation (ou la fréquence de symétrie) peut être calculé afin que les raies du spectre bande haute ainsi créé concordent au mieux avec le caractère harmonique du signal musical. Si on peut faire en sorte que cela marche bien avec un signal mono-pitch (comme la parole), il est impossible de concorder parfaitement toutes les harmoniques d'un signal polyphonique. L'oreille est cependant moins précise dans les hautes fréquences et donc cela n'est gênant que dans une moindre mesure.

Ensuite, la bande hautes fréquences ainsi construite peut être traitée afin de modeler son enveloppe spectrale. Nous discutons de cet aspect dans la prochaine partie.

2.3 Modelage de l'énergie

Une fois l'énergie créée dans les hautes fréquences, il peut être utile, voire nécessaire, de modéliser l'enveloppe spectrale dans cette bande de fréquence. Il suffit de regarder la figure 2.3 pour s'en convaincre. Dans certains formats de compression audio, tels que des dérivés de AAC, une information jointe décrit la forme de l'enveloppe spectrale dans les fréquences aiguës. Ici, nous nous intéressons aux techniques aveugles d'enrichissement spectral, c'est-à-dire que nous supposons que nous n'avons pas accès à cette information : il est donc nécessaire de prédire la forme de l'enveloppe spectrale dans les hautes fréquences. De nombreuses solutions sont proposées dans la littérature. On peut trouver deux catégories : celles qui nécessitent une étape d'apprentissage sur des signaux pleine bande, et celles qui sont fondées uniquement sur l'analyse du spectre bande basse du signal d'entrée.

2.3.1 Modèles simples d'enveloppe

Ici, il n'y a pas d'étape d'apprentissage. L'enveloppe dans la bande haute est modélisée en fonction de l'enveloppe de la bande basse. On retiendra par exemple [11] où l'on propose un modèle d'exponentielle décroissante d'enveloppe et où le paramètre de ce modèle est estimé à partir de spectre en basses fréquences.

2.3.2 Modèles statistiques

Un autre type d'approche est d'établir un modèle statistique de dépendance entre des caractéristiques bande basse et bande haute. C'est l'approche utilisée dans [10], où un modèle de mélange de gaussiennes, basé sur les coefficients cepstraux (représentant l'enveloppe spectrale) et le rapport des énergies entre les bandes basse et haute, est appris sur une base d'apprentissage de signaux pleine bande.

2.3.3 Approche par dictionnaire et quantification vectorielle

Cette technique est très bien détaillée dans [9]. Il s'agit de trouver un certain nombre de formes d'enveloppes spectrales pleine bande qui représentent au maximum l'ensemble des formes d'enveloppes que l'on peut trouver. Pour cela, on effectue un algorithme de clustering automatique (algorithme qui permet de regrouper en classes des données qui sont proches, selon une distance définie), de type k-means (voir [3]), sur les enveloppes spectrales d'une base d'apprentissage, et ensuite on choisit un représentant pour chacune des classes trouvées. Lors de l'étape de l'enrichissement spectral, on cherche le représentant qui correspond au mieux dans le bas du spectre avec le spectre bande basse à enrichir, puis on prolonge par la partie haute fréquence du représentant. La figure 2.4 illustre ce principe.

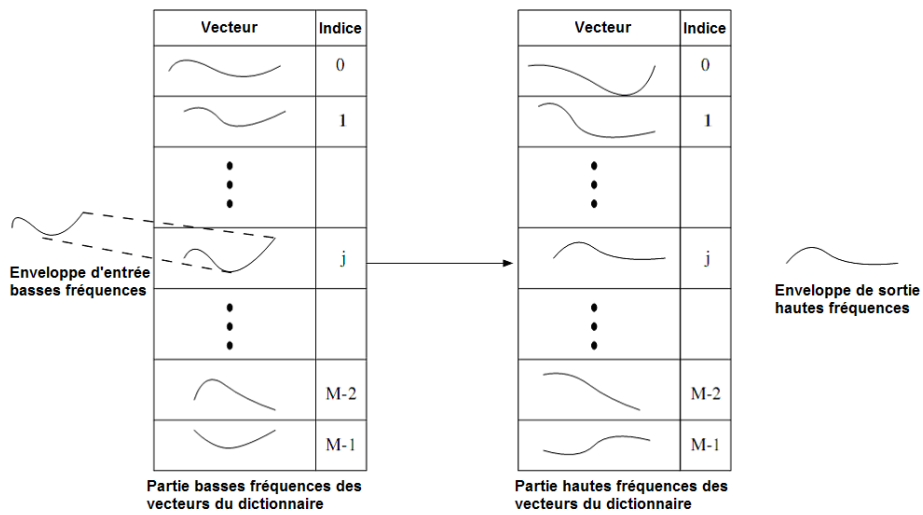


FIGURE 2.4 – Principe de l'extension de bande par quantification vectorielle

Chapitre 3

Enrichissement par décomposition atomique

Comme nous l'avons vu dans la section précédente, le schéma classique pour l'enrichissement spectral est, dans un premier temps, de créer de l'énergie dans les hautes fréquences, puis de modéliser ces hautes fréquences pour qu'elles aient une certaine enveloppe spectrale. La première approche sur laquelle nous avons travaillé diffère des approches précédentes. L'idée principale est de se servir d'un outil de décomposition atomique d'un signal audio. Dans ce chapitre, après avoir présenté l'idée sous-jacente de l'utilisation d'un tel outil, nous décrirons en détail la technique de décomposition utilisée puis présenterons un algorithme complet d'enrichissement spectral.

3.1 Principe de base

L'algorithme que nous avons mis en place ne sépare pas les deux étapes de création d'énergie et de modélisation de l'enveloppe spectrale, il fait tout en même temps. Le principe repose sur la possibilité de décomposer le spectre de puissance d'un signal comme la somme pondérée de spectres de bases (les atomes). La première étape, consistera donc à construire un dictionnaire d'atomes pleine bande qui permettront de décomposer au mieux n'importe quel signal audio (typiquement, un atome peut être un spectre harmonique de raies). Pour l'étape d'enrichissement, il suffira de décomposer le spectre d'un signal bande basse comme la somme pondérée d'atomes du dictionnaire, qu'on aura préalablement restreints à la bande basse. Nous pourrons alors élargir la bande du signal bande basse grâce aux coefficients de pondération trouvés pendant la décomposition, ainsi que la partie bande haute des atomes du dictionnaire.

Finalement, le principe est assez proche de l'approche par dictionnaire décrite à la section 2.3.3, sauf que les atomes sont des spectres de puissance (et non des enveloppes spectrales) et qu'au lieu de choisir un unique élé-

ment du dictionnaire pour un spectre bande basse donné, on cherche une combinaison linéaire d'éléments. La figure 3.1 illustre ce principe.

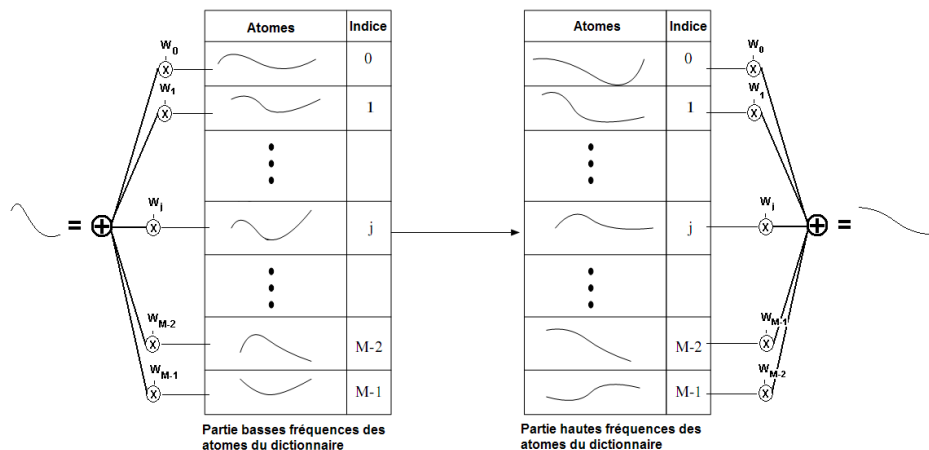


FIGURE 3.1 – Principe de l'enrichissement spectral par décomposition atomique

L'idée sous-jacente à cette méthode est d'enrichir séparément chaque source d'une mixture, à condition bien sûr que chaque atome du dictionnaire représente une source élémentaire (le spectre d'une note d'un instrument de musique par exemple).

Ayant décrit le principe général, nous allons maintenant explorer chaque détail de l'algorithme d'enrichissement, à commencer par la décomposition atomique.

3.2 Factorisation en matrices non-négatives (NMF)

Pour effectuer la décomposition en atomes d'un signal audio, la première étape est de calculer son spectrogramme de puissance. Pour cela, nous calculons le module au carré de la transformé de Fourier à court terme (TFCT) du signal, et ne gardons que les fréquences positives. Le spectrogramme de puissance est noté V . C'est une matrice $[I \times K]$ qui ne contient que des coefficients positifs. Le principe de la décomposition atomique est d'exprimer V de la manière suivante :

$$V_{i,k} = \sum_{j=1}^J W_{i,j} H_{j,k} + R_{i,k}$$

soit, en écriture matricielle :

$$V = WH + R \tag{3.1}$$

où W est une matrice $[I \times J]$ à coefficients positifs, dont chaque colonne représente un spectre de base (un atome), H une matrice $[J \times K]$ dont chaque colonne correspond aux coefficients de pondération de chaque atome de W pour un temps donné, et R le résiduel.

Lors de l'élaboration du dictionnaire, il faut trouver à la fois les matrices W et H (ce sont les colonnes de W qui seront les atomes du dictionnaire) alors que pour l'étape d'extension de bande, W est fixé (le dictionnaire) et c'est H uniquement qui nous intéresse.

La décomposition 3.1 s'appelle factorisation en matrices non-négatives, ou NMF (pour Non-negative Matrix Factorization). Le principe de l'algorithme NMF est de minimiser, selon W et H , un critère D , fonction de V , W et H :

$$(W, H) = \arg \min_{W, H} D(V, W, H)$$

Typiquement, ce critère peut être la distance euclidienne au carré, définie par :

$$D_E(V, W, H) = \sum_{i,k} (V_{i,k} - (WH)_{i,k})^2 \quad (3.2)$$

Mais il existe d'autres critères, comme la divergence de Kullback-Leibler, mieux adaptée pour des signaux audio, comme expliqué dans [12] :

$$D_{KL}(V, W, H) = \sum_{i,k} V_{i,k} \log \left(\frac{V_{i,k}}{(WH)_{i,k}} \right) - V_{i,k} + (WH)_{i,k} \quad (3.3)$$

Le critère utilisé est normalement une distance (ou du moins une mesure de divergence) mais il peut également être composé d'une distance et d'une fonction de coût destiné à imposer une contrainte sur les résultats. Utiliser une contrainte peut s'avérer utile suivant l'application que l'on fait de la NMF. Nous introduirons une contrainte spécifique de largeur spectrale pour la création du dictionnaire. Cela sera détaillé dans la section 3.3.2

Pour minimiser la mesure définie, nous pouvons effectuer une simple descente de gradient. Pour cela, il suffit de tirer aléatoirement les matrices W et H pour l'initialisation et d'appliquer de manière itérative une règle de mise à jour, définie par :

$$W \leftarrow W - \mu \nabla D_W(V, W, H) \quad H \leftarrow H - \eta \nabla D_H(V, W, H) \quad (3.4)$$

avec

$$\begin{aligned} [\nabla D_W (V, W, H)]_{i,j} &= \frac{\partial D}{\partial W_{i,j}} (V, W, H) \\ [\nabla D_H (V, W, H)]_{j,k} &= \frac{\partial D}{\partial H_{j,k}} (V, W, H) \end{aligned}$$

Ici, μ et η sont des pas suffisamment petits pour que le critère diminue et l'algorithme converge. Le problème d'une simple descente de gradient est dans la lenteur de la convergence d'une part et dans le fait que rien ne permet d'assurer la positivité des coefficients de W et H d'autre part. Pour remédier à cela, dans [8], on propose de décomposer $\nabla D_W (V, W, H)$ et $\nabla D_H (V, W, H)$ comme la différence de deux matrices à coefficients positifs,

$$\begin{aligned} \nabla D_W (V, W, H) &= \nabla^+ D_W (V, W, H) - \nabla^- D_W (V, W, H) \\ \nabla D_H (V, W, H) &= \nabla^+ D_H (V, W, H) - \nabla^- D_H (V, W, H) \end{aligned}$$

et de poser à chaque itération :

$$\mu = \frac{W}{\nabla^+ D_W (V, W, H)} \quad \eta = \frac{H}{\nabla^+ D_H (V, W, H)}$$

Les règles de mise à jour 3.4 deviennent alors :

$$W \leftarrow W \otimes \frac{\nabla^- D_W (V, W, H)}{\nabla^+ D_W (V, W, H)} \quad H \leftarrow H \otimes \frac{\nabla^- D_H (V, W, H)}{\nabla^+ D_H (V, W, H)} \quad (3.5)$$

\otimes et la barre de fraction étant respectivement la multiplication et la division, coefficients par coefficients, de deux matrices. Ici on est assuré de la positivité des matrices W et H mais comme le pas n'est plus nécessairement petit, rien ne prouve que l'algorithme converge, ni même que le critère diminue. Cependant, dans [8], il est prouvé que le critère diminue avec ces règles de mise à jour s'il correspond à la distance euclidienne ou à la divergence de Kullback-Leibler. C'est l'algorithme NMF multiplicatif.

3.3 Apprentissage du dictionnaire

Maintenant que nous avons présenté l'outil de la factorisation en matrices non-négatives, nous pouvons détailler le processus d'apprentissage du dictionnaire. Le principe est de faire tourner l'algorithme NMF sur un signal audio, et d'extraire de la matrice W ainsi trouvée les atomes du dictionnaire. Nous retiendrons deux possibilités : ou bien laisser à l'algorithme NMF le bon soin de trouver les atomes du dictionnaire, en le faisant tourner sur la concaténation de plusieurs extraits musicaux variés, ou alors appliquer l'algorithme plusieurs fois sur de courts extraits musicaux bien particulier, par exemple des notes isolées d'instruments.

3.3.1 Apprentissage non supervisé

La première possibilité a l'avantage d'être un apprentissage non supervisé, mais hélas, elle ne s'avère pas très pertinente. Tout d'abord, la complexité de l'algorithme NMF n'est pas linéaire en la taille des données, et il n'est pas possible, aussi bien en terme de temps de calcul, qu'en terme de mémoire vive, d'appliquer l'algorithme sur un trop long fichier audio. D'autre part, la NMF est un bon outil pour réduire la dimension de l'espace (ici le nombre I de lignes de la matrice V), c'est-à-dire trouver une matrice W ayant plus de lignes que de colonnes, mais n'est pas très adaptée pour créer une famille redondante (ou même une base vectorielle) d'atomes. En effet, supposons que nous cherchions une décomposition NMF avec W une matrice carrée et la distance euclidienne comme mesure critère de minimisation. Alors la NMF a de forte chances de converger vers $W = I_d$ et $H = V$.

3.3.2 Apprentissage supervisé

Nous avons donc choisi de diriger nos recherches vers la deuxième possibilité, même si, pour cela, il est nécessaire de créer une large base de données avec des notes séparées de chaque instrument. Chaque fichier audio est décomposé, selon l'algorithme NMF, en un petit nombre fixé d'atomes (typiquement entre deux et quatre). En ce qui concerne le critère de convergence, il est composé de la divergence de Kullback-Leibler, ainsi que d'une contrainte sur la largeur de bande de chaque atome (les colonnes de W). En effet, nous voulons nous assurer que les atomes trouvés sont des atomes pleines bandes car si leur énergie était concentrée sur une faible largeur de bande, alors ces atomes ne pourrait pas être utilisé pour l'enrichissement. Dans [4], on donne un critère de largeur spectrale défini par :

$$L(W) = \sum_j \left(\frac{\sum_i f_i^2 W_{i,j}}{\sum_i W_{i,j}} - \left(\frac{\sum_i f_i W_{i,j}}{\sum_i W_{i,j}} \right)^2 \right)$$

où f_i est la fréquence réduite correspondant au bin fréquentiel i . Plus $L(W)$ est grand, plus chaque colonne de W est à bande large.

Ainsi, le critère de minimisation est défini comme :

$$D(V, W, H) = (1 - \lambda)D_{KL}(V, W, H) - \lambda L(W) \quad (3.6)$$

où $D_{KL}(V, W, H)$ est la divergence de Kullback-Leibler donnée par la formule 3.3 et λ un coefficient compris dans l'intervalle $[0, 1]$ qui tient compte de l'importance de la contrainte de largeur spectrale.

On peut calculer les dérivées partielles de $D(V, W, H)$ et utiliser les règles de mise à jour 3.5 avec :

$$\begin{aligned} [\nabla^- D_W (V, W, H)]_{i,j} &= \sum_k \frac{H_{j,k} V_{i,k}}{(WH)_{i,k}} + \frac{f_i^2 + 2 \frac{(\sum_m f_m W_{m,j})^2}{(\sum_m W_{m,j})^2}}{\sum_m W_{m,j}} \\ [\nabla^+ D_W (V, W, H)]_{i,j} &= \sum_p H_{j,p} + \frac{\sum_m W_{m,j} (f_m^2 + 2 f_i f_m)}{(\sum_m W_{m,j})^2} \\ [\nabla^- D_H (V, W, H)]_{j,k} &= \sum_m \frac{W_{m,j} V_{m,k}}{(WH)_{m,k}} \\ [\nabla^+ D_H (V, W, H)]_{j,k} &= \sum_m W_{m,j} \end{aligned}$$

Rien ne prouve qu'avec ces règles de mise à jours l'algorithme converge, ni même que le critère diminue, mais on a constaté la diminution systématique du critère tant que le paramètre λ n'est pas trop élevé. La figure 3.2 permet d'illustrer le résultat de la NMF avec et sans la contrainte de largeur spectrale. Sur cette figure, nous pouvons voir tout d'abord un spectrogramme composé de deux spectres (deux colonnes) : un spectre constant égal à 1 sur toute la bande de fréquence (première colonne) et un spectre constant égal à 1 sur seulement la moitié basse du spectre et 0 ailleurs (deuxième colonne). Ce genre de spectrogramme permet par exemple de décrire grossièrement l'allure du spectrogramme d'une note de piano, où les harmoniques aiguës s'estompent avant les harmoniques graves. Nous avons décomposé ce spectrogramme en deux atomes.

Dans la pratique, étant donnée le spectrogramme d'un signal audio de la base d'apprentissage, l'algorithme NMF avec contrainte de largeur spectrale est effectué avec un nombre d'itération fixe (par exemple 200 itérations) et un facteur λ petit (typiquement $\lambda = 0.1$)

3.4 Enrichissement

Une fois le dictionnaire appris, la phase d'enrichissement est plutôt simple. Si V est le spectrogramme d'amplitude d'un signal audio bande basse à enrichir, alors il peut être décomposé de la forme :

$$V = \begin{bmatrix} V_{bb} \\ V_{bh} \end{bmatrix}$$

où les coefficients de la matrice V_{bh} sont quasi nuls. De plus, si W est le dictionnaire (chaque colonne est un atome), alors il est décomposable de la même manière,

$$W = \begin{bmatrix} W_{bb} \\ W_{bh} \end{bmatrix}$$

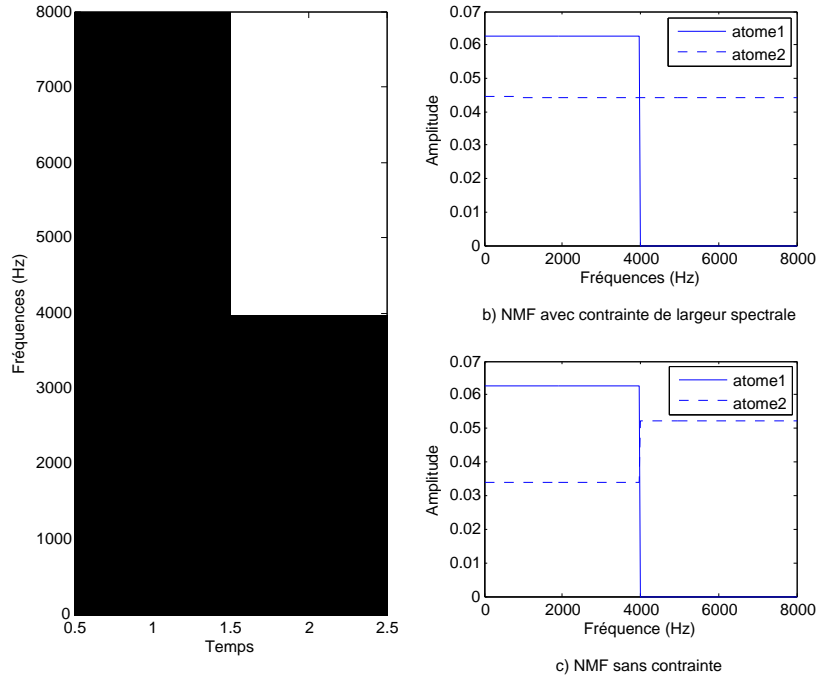


FIGURE 3.2 – Décomposition d’un spectrogramme (a) en deux atomes en utilisant la NMF avec (b) et sans (c) contrainte de largeur spectrale

L’algorithme d’enrichissement consiste alors à trouver, grâce à l’algorithme NMF, H tel que :

$$H = \arg \min_H D_{KL}(V_{bb}, W_{bb}, H)$$

Pour cela, il suffit, pour chaque itération de l’algorithme NMF, de mettre à jour uniquement la matrice H avec la règle 3.5. On a donc décomposé la partie basse des spectres à enrichir comme la somme pondérée d’atomes du dictionnaire (restreints en bande basse). On peut alors utiliser la partie hautes fréquences des atomes du dictionnaire pour l’enrichissement, en posant :

$$V'_{bh} = W_{bh}H;$$

Alors le nouveau spectrogramme d’amplitude du signal enrichi est :

$$V' = \begin{bmatrix} V_{bb} \\ V'_{bh} \end{bmatrix}$$

Il reste une dernière difficulté pour reconstruire la forme d’onde du signal enrichi : il s’agit du fait que V' est un spectrogramme en amplitude et, si nous pouvons lors du calcul initial de V sauvegarder les phases de V_{bb} , il nous

manque bel et bien les phases de V'_{bh} . Cependant dans [13], il est proposé un algorithme pour reconstruire les phases, et ainsi utiliser la TFCT inverse, algorithme que nous avons utilisé.

3.5 Discussion

La méthode que nous avons mise en place est une méthode originale, en rupture avec les approches “classiques” décrites dans le chapitre 2. Elle pose cependant un problème de taille : l’infinie richesse des signaux musicaux implique que le nombre d’atomes du dictionnaire doit être très élevé. Les tests que nous avons effectués ont été uniquement fait sur des morceaux de piano et seront dépeints dans le chapitre 5.

Chapitre 4

Enrichissement du bruit par modèle de signal

Une autre approche, complètement différente de la précédente, a été également élaborée pendant ce stage. Elle est fondée sur un modèle mathématique de signal audio et a la particularité d'enrichir séparément la partie aléatoire (le bruit) de la partie déterministe (les sinusoïdes). L'idée vient de l'observation que pour de nombreux morceaux de musique (la musique pop par exemple), le bruit dans les hautes fréquences est largement prépondérant. Dans ce chapitre, nous nous intéressons uniquement à l'enrichissement du bruit.

4.1 Modèle de signal

Le modèle de signal que nous avons retenu est un modèle assez courant qui a l'avantage de donner une assez bonne description des signaux musicaux. Il s'agit d'un modèle de somme de sinusoïdes plus du bruit coloré :

$$x(n) = \sum_{l=1}^L A_l e^{2i\pi f_l n} + (b \otimes e)(n) \quad (4.1)$$

où les A_l sont des amplitudes complexes, f_l des fréquences réduites, $b(n)$ un bruit blanc centré gaussien de variance σ^2 et e un filtre lentement variable, rendant compte de la coloration du bruit.

En prenant le module au carré de la transformé de Fourier, et en estimant le module d'une somme par la somme des modules (cela est bien évidemment faux mais on peut cependant prouver que l'espérance du module au carré d'une somme de deux complexes ayant des phases aléatoires uniformes sur $[0, 2\pi]$ et indépendantes est égale à la somme des modules au carré), on

obtient :

$$V(i) = |X(i)|^2 \approx \sum_{l=1}^L C_l P_l(i) + |B(i)|^2 |E(i)|^2 \quad (4.2)$$

où les $P_l(i)$ sont les spectres de puissance de chaque sinusoïde (pics centrés en f_l), les C_l les amplitudes de ces pics, $|B(i)|^2$ est un bruit blanc de loi exponentielle de paramètre $1/\sigma^2$, et où $|E(i)|^2$ correspond à l'enveloppe spectrale du bruit coloré.

Le principe de l'enrichissement spectral du bruit est dans un premier temps de créer un dictionnaire d'enveloppes pleine bande. Pour cela, le spectrogramme de puissance V d'un signal d'apprentissage est calculé, puis les enveloppes spectrales du bruit sont estimées pour chaque colonne de V et enfin une étape de classification automatique permet de garder un nombre restreint d'enveloppes représentatives qui constitueront le dictionnaire. Une fois le dictionnaire créé, le spectrogramme d'un signal bande basse à enrichir est calculé et les enveloppes spectrales estimées de la même manière. Pour étendre la bande, il suffira, pour chaque enveloppe estimée, de chercher dans le dictionnaire l'enveloppe qui correspond le mieux en bande basse et de l'utiliser pour enrichir les aiguës avec du bruit coloré.

4.2 Algorithme d'enrichissement du bruit

4.2.1 Extraction des enveloppes

Que ce soit pour la création du dictionnaire ou l'étape d'enrichissement, il faut estimer l'enveloppe spectrale du bruit coloré d'un court échantillon de musique. Supposons que l'on ait un spectre de puissance $V_k(i)$ tel que défini par le modèle 4.2 (l'indice k signifie qu'il correspond à une colonne d'un spectrogramme de puissance V). Pour estimer l'enveloppe spectrale du bruit, il faut donc à la fois éliminer les pics dus à la présence de sinusoïdes, et lisser le spectre.

Pour cela, il est possible d'utiliser un filtre de rang. La longueur q du filtre de rang a été choisie empiriquement comme étant égale au cinquième du nombre de points sur lesquels ont été calculés la transformée de Fourier discrète. Pour calculer la valeur du spectre lissé en chaque point, les q valeurs extraites sont triées par ordre croissant, puis celle d'ordre $q/3$ est sélectionnée (cette valeur a été également choisie de manière empirique). La figure 4.1 illustre cette étape.

Une fois le filtre de rang appliqué, nous modélisons l'enveloppe E_j obtenue comme une somme pondérée de fenêtres de Hann, centrés sur des fréquences

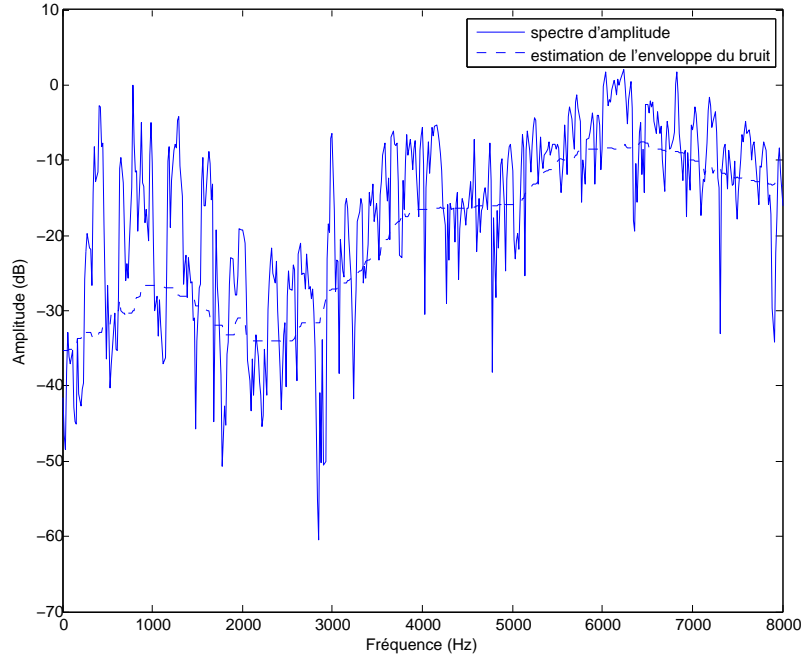


FIGURE 4.1 – Estimation de l’enveloppe spectrale du bruit grâce à un filtre d’ordre

réduites f_j régulièrement espacées entre 0 et 0.5 :

$$E_k(i) = \sum_{j=1}^J P_k(j)W_{f_j}(i) + R_k(i)$$

où $W_{f_j}(k)$ est un spectre nul sauf autour de la fréquence f_j où il est égal à une fenêtre de Hann, où les $P_k(j)$ sont des coefficients positifs et où $R_k(i)$ est le résiduel. En écriture matricielle, cela donne :

$$E_k = WP_k + R_k \quad (4.3)$$

La largeur des fenêtres est calculée de sorte qu’elle corresponde à une largeur spectrale de 1500 Hz et il y a un recouvrement de 75 % entre deux fenêtres consécutives (ce qui permet de déduire l’espacement entre les f_i). Le type de fenêtre (Hann) et le recouvrement ont été choisis de sorte que si tout les coefficients sont égaux à 1, alors la somme pondérée vaut une constante. L’enveloppe est maintenant décrite par le vecteur P des poids des fenêtres de Hann. Cette étape permet à la fois de lisser l’enveloppe spectrale du bruit et de réduire la dimension de l’espace des paramètres d’enveloppe. Le vecteur P est trouvé en appliquant l’algorithme NMF (avec la divergence de Kullback-Leibler). La matrice W reste fixe et seul le vecteur P est mis à jour. Sur

la figure 4.2, on peut voir une enveloppe E estimée grâce au filtre d'ordre, l'ensemble des fenêtres de hann pondérées par les coefficients de P ainsi que la nouvelle enveloppe $E' = WP$.

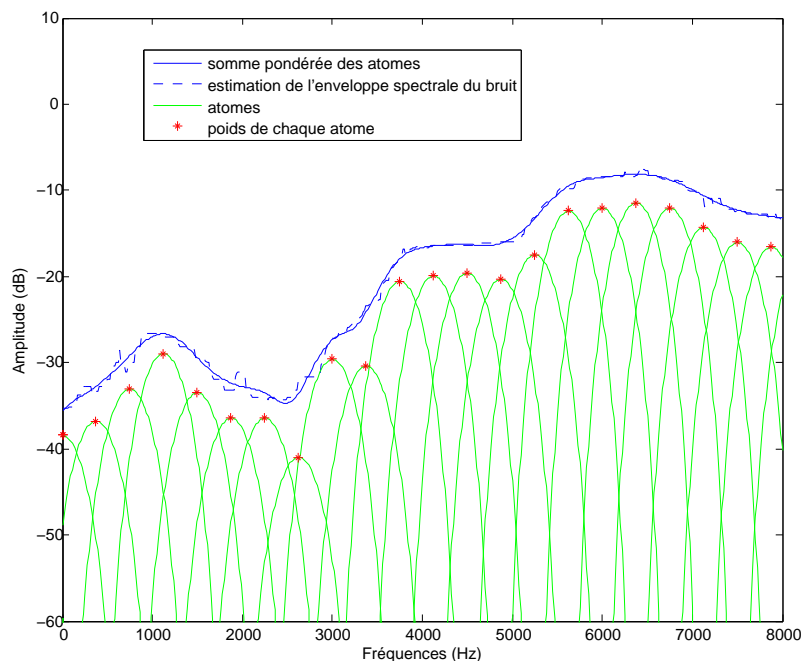


FIGURE 4.2 – Modélisation d’une enveloppe spectrale, comme la somme pondérée de fenêtres de Hann

4.2.2 Création du dictionnaire

Pour construire le dictionnaire, il faut d’abord créer une base d’apprentissage. Pour cela nous avons pris plusieurs extraits de musique de tout genre. Puis nous avons calculé le spectrogramme V de la concaténation de tous les extraits, avec des fenêtres de Hamming dont la taille sera discuter dans le chapitre 5, avec un recouvrement entre deux fenêtres de 75 %. Pour chaque colonne V_k du spectrogramme, l’enveloppe spectrale E_k du bruit est estimée puis P_k est calculé selon l’équation 4.3 grâce à la NMF.

Une fois la matrice P calculée, il est nécessaire de choisir un nombre réduit de représentants pour construire le dictionnaire. Pour cela, un algorithme des k-means ([3]) est utilisé sur l’ensemble des colonnes de P et seuls les représentants de chaque classe sont ajoutés au dictionnaire. Comme nous voulons un dictionnaire de formes d’enveloppe (sans prendre en compte la puissance de l’enveloppe spectrale), nous avons défini une distance particulière pour l’algorithme des k-means. Cette distance est inspirée de la distance log-spectrale mais elle doit être invariante par un facteur multiplicatif. Elle

est définie comme :

$$D(P_{k_1}, P_{k_2}) = \min_c \|\log(P_{k_1}) - \log(P_{k_2}) - c\|^2 \quad (4.4)$$

où P_{k_1} et P_{k_2} sont deux colonnes différentes de la matrice P . En dérivant la fonction D par rapport à c , et en cherchant le minimum, on obtient :

$$D(P_{k_1}, P_{k_2}) = \left\| \log(P_{k_1}) - \log(P_{k_2}) - \frac{1}{J} \sum_{j=1}^J (\log(P_{k_1}(j)) - \log(P_{k_2}(j))) \right\|^2 \quad (4.5)$$

où J est le nombre de ligne de la matrice P . Ainsi, $D(P_{k_1}, P_{k_2}) = 0$ si $P_{k_2} = aP_{k_1}$, a étant un réel.

Sur la figure 4.3 est dessiné en dB l'ensemble des enveloppes spectrales d'une classe après l'algorithme des k-means. On remarque bien que leurs formes se ressemblent, même si elles ont un niveau d'énergie tous différents.

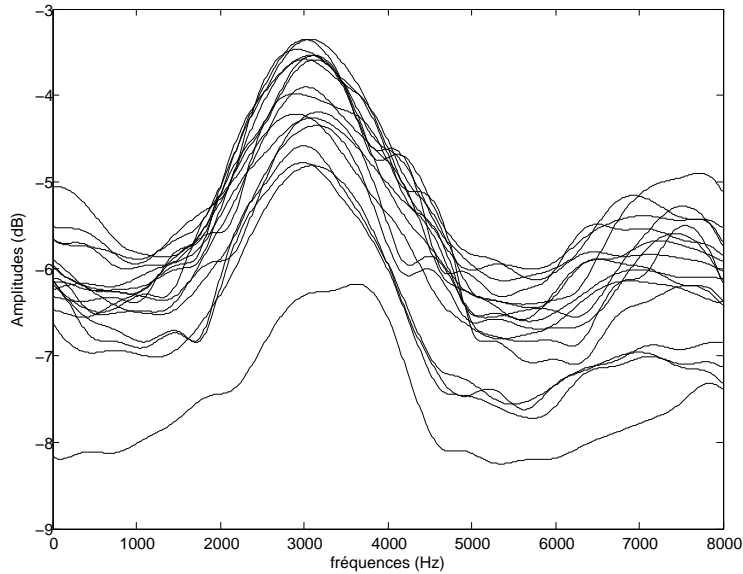


FIGURE 4.3 – Ensemble de toutes les enveloppes spectrales appartenant à une même classe

4.2.3 Enrichissement

L'étape d'enrichissement est la même que celle décrite dans la section 2.3.3. Etant donnée un signal bande basse, on calcule son spectrogramme V

de la même manière que pour l'étape de création du dictionnaire, on en déduit les enveloppes spectrales E du bruit ainsi que la matrice P des coefficients du modèle d'enveloppe. Ces coefficients sont tous quasi nuls à partir de la fréquence de coupure du signal bande basse et donc P peut s'écrire sous la forme :

$$P = \begin{bmatrix} P_{bb} \\ P_{bh} \end{bmatrix}$$

De la même manière, le dictionnaire B peut alors s'écrire :

$$B = \begin{bmatrix} B_{bb} \\ B_{bh} \end{bmatrix}$$

Pour chaque colonne de P_{bb} , on cherche dans B_{bb} , la colonne qui minimise la distance 4.5, et on construit ainsi P'_{bh} en y ajoutant la colonne de B_{bb} correspondante, multipliée par le bon facteur multiplicatif. Ainsi pour chaque colonne k de P_{bb} , on pose :

$$p_0 = \arg \min_p D(P_{bb}(k), B_{bb}(p))$$

et :

$$P'_{bh}(k) = \exp \left(\sum_{j=1} (\log(P_{bbj,k}) - \log(B_{bbj,p_0})) \right) B_{bh}(p_0)$$

En posant :

$$P' = \begin{bmatrix} P_{bb} \\ P'_{bh} \end{bmatrix}$$

on peut calculer les nouvelles enveloppes spectrales du bruit du signal bande basse enrichi.

$$E' = WP'$$

Avant de pouvoir reconstruire le spectre de puissance étendu V' en tirant aléatoirement des échantillons selon une loi exponentielle, il faut encore estimer le paramètre de la loi pour chaque colonne de V' . Pour cela, nous construisons le spectrogramme W comme :

$$W = \frac{V}{E'}$$

où la barre de fraction correspond à une division coefficients par coefficients. W correspond donc au spectrogramme de notre signal bande basse, où l'on

aurait blanchi le bruit. Comme les coefficients de V , et à fortiori de W sont tous quasi nuls à partir d'une fréquence de coupure on peut écrire :

$$W = \begin{bmatrix} W_{bb} \\ W_{bh} \end{bmatrix}$$

On peut alors estimer les paramètres de la loi exponentielle du bruit pour chaque colonne grâce à la matrice W_{bb} . En effet, chaque colonne k de cette matrice suit le modèle :

$$W_{bb}(k) \approx \sum_{l=1}^L C_l P_l(i) + |B(i)|^2$$

et la médiane λ_k de $W_{bb}(k)$ semble être un bon estimateur du paramètre.

Pour construire le spectrogramme de puissance enrichi V' du signal basse bande enrichi, il faut donc dans un premier temps estimer les paramètres λ_k , puis construire V'_{bh} telle que pour chaque colonne k , on ait :

$$V'_{bh}(k) = B_{\lambda_k} \otimes E'_{bh}$$

où B_{λ_k} est un vecteur aléatoire dont chaque coefficient est tiré selon une loi exponentielle de paramètre λ_k et \otimes la multiplication termes par termes. V' s'écrit alors :

$$V' = \begin{bmatrix} V_{bb} \\ V'_{bh} \end{bmatrix}$$

Pour construire la forme d'onde du signal enrichi, on utilise aussi l'algorithme développé dans [13].

Chapitre 5

Tests et résultats préliminaires

Dans cette partie, nous allons discuter des résultats obtenus après avoir testé chacun des algorithmes conçus pendant ce stage. Pour la reconstruction des hautes fréquences, il est difficile de mesurer une amélioration perceptible de la qualité audio. L'idée d'origine était de vérifier l'amélioration de la qualité audio en comparant un signal pleine bande original avec, d'une part le même signal bande basse, et d'autre part le signal bande basse enrichi grâce à nos algorithmes. Pour cela, nous aurions utilisé le système PEMO-Q [6], qui permet d'établir une distance perceptible entre deux signaux audio. Si la distance trouvée entre le signal original et le signal reconstruit est plus faible qu'entre le signal original et le signal bande basse, cela signifie que nos algorithmes sont prometteurs. Les tests auraient été faits sur une grande base de données, contenant des genres musicaux différents, et les résultats auraient été classés suivant le genre, afin d'établir à quels types de musique nos systèmes d'enrichissement spectral s'appliquent le mieux.

Hélas, nous n'avons pas eu le temps d'effectuer ces tests perceptifs et nous n'avons donc pas encore de résultats théoriques. Ils seront faits prochainement. En attendant, nous avons quand même fait tourner nos algorithmes sur quelques exemples musicaux, nous les avons écoutés, analysés et nous avons identifié un certain nombre de problèmes et de résultats positifs.

5.1 Enrichissement par décomposition atomique

Dans cette section, nous présentons les résultats de l'algorithme décrit dans le chapitre 3.

5.1.1 Mise en place du test

Comme nous l'avons annoncé dans la section 3.5, l'algorithme d'enrichissement par décomposition atomique nécessite d'apprendre les atomes de bases à partir de notes isolés d'instruments spécifiques. Comme il n'est

pas possible que créer un dictionnaire trop grand (pour des raisons de complexité algorithmique), nous avons appris le dictionnaire sur des notes isolés de piano, et, a fortiori, testé l’algorithme sur des morceaux de piano solo.

Ainsi, la base d’apprentissage comprends :

- 88 fichiers comprenant l’enregistrement de 88 notes *forte* d’un piano,
- 88 fichiers comprenant les mêmes notes *mezzo forte*,
- 88 fichiers comprenant les notes *piano*.

Ces fichiers sont des fichiers au format *wav* échantillonnés à 16000 Hz et sont extraits de la base de données RWC [5].

Pour chacun de ces fichiers, nous avons d’abord calculé de spectrogramme de puissance grâce à la TFCT en utilisant des fenêtres de Hamming de 1024 échantillons et un recouvrement de 75 %. Pour chacun des spectrogrammes, nous appris 2 atomes de base en utilisant la NMF avec contrainte de largeur spectrale (section 3.3.2), avec 200 itérations et un paramètre λ (équation 3.6) égal à 0.1. Le dictionnaire est donc composé de 524 atomes. Sur la figure 5.1 est représenté le spectrogramme d’une note de piano, la reconstruction du spectrogramme selon la NMF, ainsi que les deux atomes extraits. Une fois

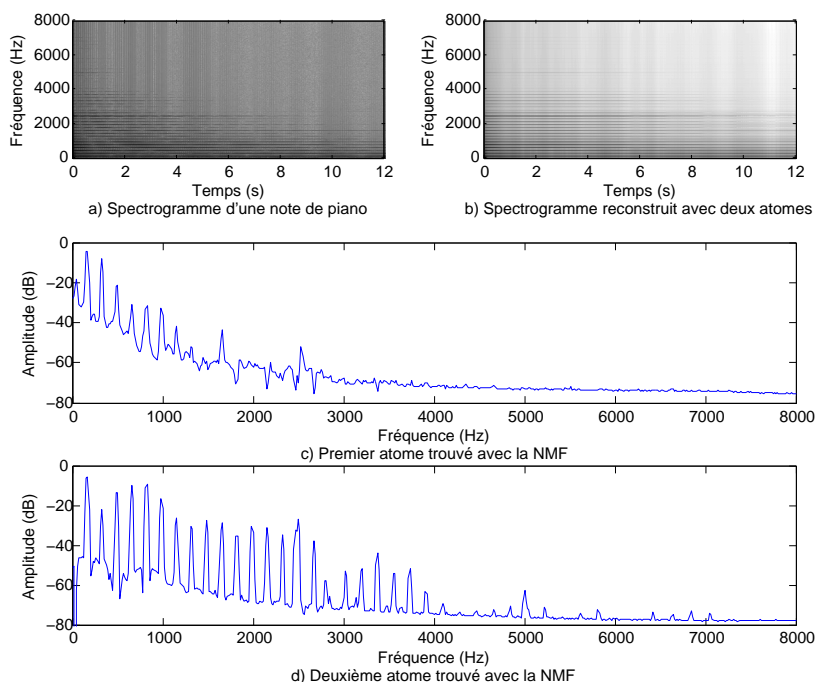


FIGURE 5.1 – Exemple de décomposition d’une note de piano en deux atomes grâce à la NMF (avec contrainte de largeur spectrale)

le dictionnaire appris, nous avons testé la méthode d’enrichissement sur plusieurs morceaux de piano, échantillonnés à 8000 Hz, chaque morceau ayant

des nuances différentes. Sur la figure 5.2, on peut voir le spectrogramme d'un signal pleine bande, ainsi que celui du signal basse bande enrichi correspondant.

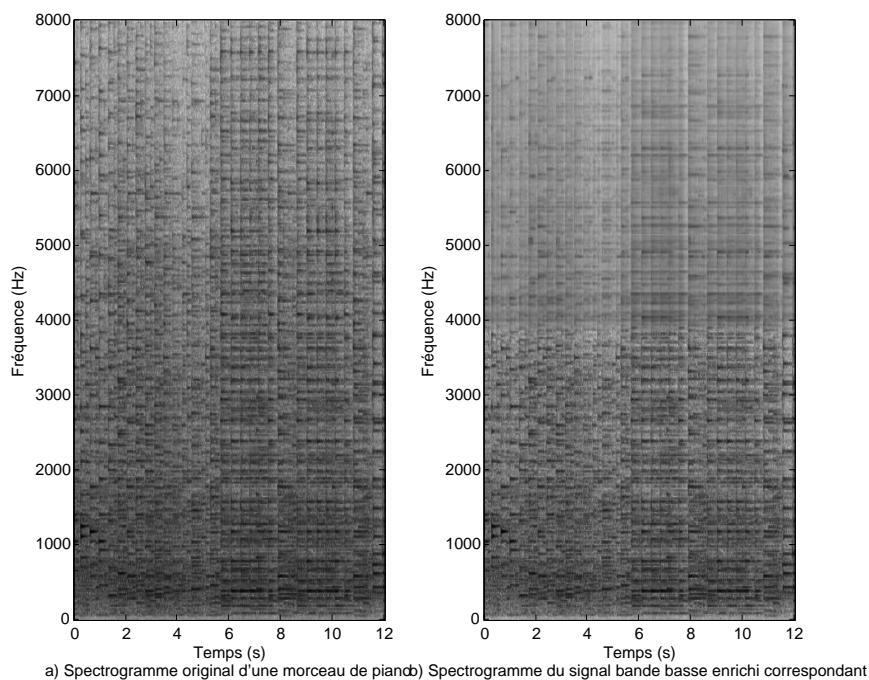


FIGURE 5.2 – Exemple d'enrichissement spectrale par décomposition atomique

5.1.2 Résultats et discussions

Nous pointons ici quelques problèmes et points positifs remarqués lors de l'écoute des signaux enrichis.

Non-discrimination des nuances Le spectre d'une note de piano est d'autant plus étendu dans les hautes fréquences que la note a été jouée *forte*. De plus, les harmoniques d'une note tenue s'estompent d'autant plus rapidement qu'elles sont d'ordres élevées. Ces effets jouent beaucoup dans la perception des nuances pour le piano. Hélas, lorsqu'on calcule le signal bande basse, alors les spectres d'une même note jouée *forte* et *piano* se ressemblent beaucoup. Il devient alors impossible pour la NMF de prédire sur quels atomes (ceux extraits des notes *forte*, *mezzo forte* ou *piano*) il faut décomposer le spectrogramme bande basse.

Il en résulte dans l'écoute d'un signal enrichi, une certaine irrégularité dans les hautes fréquences : certaines notes seront enrichies grâce à des

atomes de notes *forte*, alors que d'autres grâce à des atomes de notes *piano*, cela de manière aléatoire. Ce défaut devrait s'atténuer en augmentant la largeur de la bande basse mais nous n'avons pas encore pu essayer.

Une solution à ce problème, pourrait être de classer les atomes du dictionnaire suivant leur niveau de puissance dans la base d'apprentissage, puis de discriminer tel ou tel atome lors de la décomposition du spectrogramme bande basse pendant la NMF, suivant le niveau de puissance du spectrogramme.

Présence de fréquences gênantes Tout le principe de l'enrichissement par décomposition atomique est d'enrichir chaque note séparément et ainsi s'assurer que le spectre bande haute construit possède les mêmes propriétés harmoniques que le spectre bande basse à enrichir. Hélas, nous pouvons quand même entendre dans le signal enrichi des fréquences aiguës qui ne devraient pas être présentes. Cela est dû au fait que lors de la décomposition atomique du spectre bande basse, d'autres atomes du dictionnaire que ceux correspondants aux notes présentes dans le signal sont activés. Par exemple, un atome correspond à une note aiguë dont la fréquence fondamentale correspond à une harmonique d'une note du signal bande basse peut être activé par l'algorithme NMF lors de l'étape d'enrichissement, créant ainsi en haute fréquences de l'énergie non voulue au niveau des harmoniques de l'atome.

Pour remédier à ce problème, une idée serait d'ajouter une contrainte de *sparseness* sur les colonnes de la matrice H lors de la décomposition du signal bande basse. Autrement dit, une contrainte pour activer le moins d'atomes possible lors de l'étape d'enrichissement. Cela pourrait assurer que seules les atomes correspondants aux notes présentes dans le signal bande basse sont activés. Dans [12], une telle contrainte est décrite ainsi que les règles de mise à jour correspondantes pour l'algorithme NMF.

Brillance La caractéristique des signaux dont les fréquences aiguës ont été coupées, est un son mat, qui manque de brillance. Après enrichissement, malgré les problèmes expliqués ci-dessus, le son devient plus brillant, ce qui a pour conséquence d'augmenter le confort d'écoute.

5.2 Enrichissement du bruit par modèle de signal

L'algorithme d'enrichissement du bruit par modèle de signal introduit dans le chapitre 4 a l'avantage, contrairement à l'algorithme par décomposition atomique, d'être plus souple dans ses applications. Ainsi, les tests peuvent se faire sur n'importe quel genre de musique.

5.2.1 Mise en place du test

Le dictionnaire est appris grâce à une base de données de 15 minutes de musique échantillonnée à 16000 Hz, composé d'extraits de 15 secondes de musique de tout genre (pop, rap, électro, jazz, musique classique). La taille des fenêtres de Hamming grâce auxquelles on calcule le spectrogramme de puissance sera discutée dans la section suivante. Nous avons ensuite construit un dictionnaire composé de 2048 modèles d'enveloppes spectrales (voir section 4.2.1).

Nous avons ensuite testé l'algorithme d'enrichissement du bruit sur des morceaux de musique de plusieurs genres, échantillonnés à 8000 Hz. Sur la figure 5.3, nous pouvons voir le spectrogramme de puissance d'un signal original pleine bande, ainsi que le spectrogramme du signal bande basse enrichi correspondant.

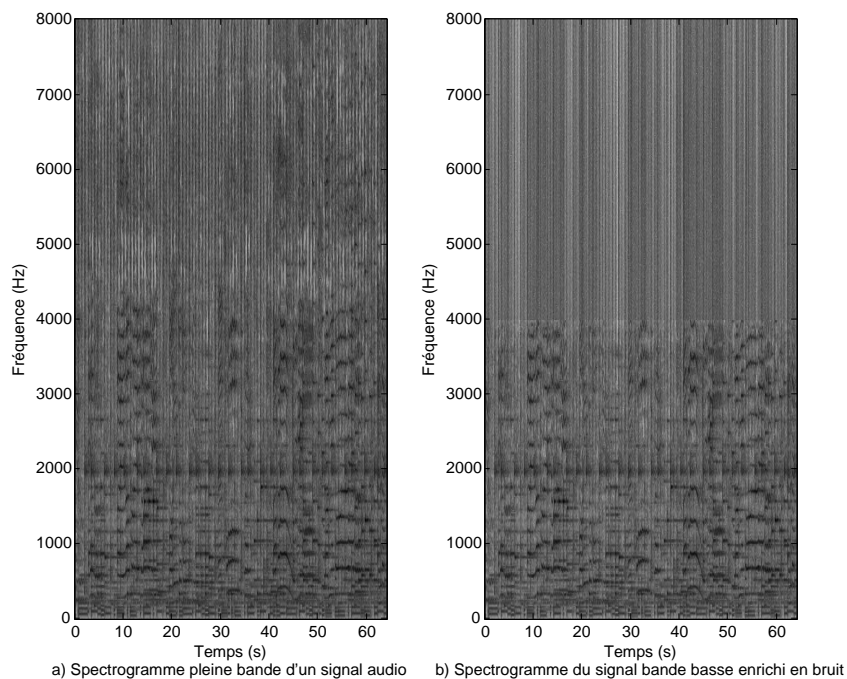


FIGURE 5.3 – Exemple d'enrichissement du bruit

5.2.2 Résultats et discussion

Influence de la taille des fenêtres d'analyse Indépendamment du style de musique sur lequel l'algorithme a été testé, la taille des fenêtres de Hamming lors du calcul des spectrogrammes de puissance a une grande influence sur les résultats. En effet, suivant la taille des fenêtres, nous favorisons la précision spectrale ou temporelle.

Si la taille de la fenêtre est trop petite, il devient difficile d'estimer l'enveloppe spectrale du bruit, car alors on ne peut plus discriminer les sinusoides présentes dans le signal : dans le domaine fréquentiel, les pics de chaque sinusoides se chevauchent et l'étape du filtrage du spectre avec un filtre d'ordre devient obsolète. Cela se traduit par une mauvaise estimation de l'enveloppe du bruit, ainsi qu'une mauvaise estimation de sa puissance. Lors du processus d'enrichissement, on ajoute alors du bruit là où il n'est pas censé y en avoir, et cela gêne l'écoute du signal enrichi.

D'autre part, si les fenêtres d'analyse sont trop grandes, alors l'estimation de l'enveloppe spectrale est de bonne qualité (les pics des sinusoides sont suffisamment espacés pour que le filtre d'ordre sur le spectre soit efficace) mais alors nous avons une mauvaise précision temporelle. Dans les tests d'écoute, cela se traduit par un pré-écho lors des attaques percussives. Autrement dit, on entend le bruit dans les hautes fréquences quelques millisecondes avant une attaque. De plus le signal enrichi manque de dynamique dans les hautes fréquences.

Le meilleur compromis que nous avons trouvé est d'utiliser des fenêtres de 512 échantillons, même si cela n'est quand même pas parfait. Un travail qu'il reste encore à fournir est donc d'estimer correctement l'enveloppe spectrale du bruit tout en gardant une bonne précision temporelle.

Qualité des résultats par style de musique Sans surprise, ce système d'enrichissement spectral est d'autant plus efficace que le signal à enrichir comporte des percussions. Aussi, c'est pour la musique pop, le jazz, ou autres musiques ayant des percussions, que l'enrichissement est de meilleure qualité. Pour la musique classique, on remarque quand même une amélioration dans les attaques percussives, comme les attaques du piano par exemple.

En revanche, plus le signal est polyphonique, plus du bruit parasite est créé. Cela est dû à la mauvaise estimation de l'enveloppe spectrale du bruit, et de son intensité, que nous avons relevée dans le paragraphe précédent : s'il y a trop de sinusoides, alors on ne peut plus discriminer les pics, ceux-ci seront considérés comme du bruit et a fortiori, du bruit sera rajouté dans les hautes fréquences alors qu'il n'y en a pas dans le signal pleine bande d'origine.

Chapitre 6

Conclusion

6.1 Résumé du rapport et des contributions apportées

Dans ce rapport, nous avons présenté les techniques d'enrichissement spectral en bande haute, ainsi que deux nouvelles approches pour ce problème.

Après une introduction sur les enjeux de l'extension de bande dans le chapitre 1, nous avons étudié dans le chapitre 2 l'état de l'art du problème d'enrichissement spectral. Nous avons remarqué que dans les approches classiques, on divise le processus d'enrichissement en deux temps. Un premier temps où l'on crée de l'énergie dans les hautes fréquences, en fonction du signal bande basse à enrichir et un deuxième temps où l'on modèle cette énergie pour obtenir une certaine enveloppe spectrale. Enfin nous avons tiré de la littérature quelques exemples représentatifs des solutions à ces deux problèmes.

Le chapitre 3 introduit un nouveau principe d'enrichissement spectral par décomposition harmonique. L'algorithme NMF est utilisé pour décomposer un spectre en une somme pondérée de spectres de base. Dans une première étape d'apprentissage, on crée un dictionnaire d'atomes pleine bande, et ces atomes sont ensuite utilisés pour décomposer un signal bande basse et ainsi étendre sa bande. Nous avons introduit une nouvelle contrainte de largeur spectrale pour l'algorithme NMF, ainsi que les nouvelles règles de mise à jour avec cette contrainte. Enfin nous avons remarqué qu'avec l'utilisation d'une telle technique, il faut un dictionnaire extrêmement grand afin de représenter toute la richesse des signaux musicaux. Cette technique d'enrichissement s'appliquerait donc plus à une utilisation supervisée, où par exemple on sait a priori quels instruments sont présents dans le signal, et on charge les dictionnaires spécifiques à ces instruments pour enrichir le signal.

Une autre méthode d'enrichissement fondée sur la séparation du bruit et des sinusoïdes, puis l'enrichissement du bruit, a été introduite dans le chapitre 4. Dans cette partie, nous avons modélisé le spectre de puissance d'un signal comme une somme de pics et du bruit coloré, nous avons conçu une méthode pour estimer l'enveloppe spectrale du bruit (partie 4.2.1), ainsi que sa puissance, et nous avons introduit une nouvelle distance entre deux enveloppes spectrales (équation 4.5) qui permet d'évaluer une distance entre formes d'enveloppes, indépendamment de leur puissance. Ensuite nous avons utilisé une technique classique de quantification vectorielle avec dictionnaire pour le processus d'enrichissement.

Enfin, nous avons dans le chapitre 5 effectué des tests préliminaires et mis le doigt sur un certain nombre de problèmes et de points positifs pour chacune des deux méthodes mises en place.

Travail futur La première chose à faire est d'effectuer les tests selon le système proposé au début du chapitre 5 afin d'avoir une mesure objective de l'amélioration de la qualité audio. Si dans ce rapport, nous nous sommes restreint à une bande basse de 0 à 4000 Hz et une pleine bande de 0 à 8000 Hz, les algorithmes que nous avons mis en place peuvent tout à fait s'appliquer pour d'autres valeurs de ces bandes. Les tests pourront donc se faire pour plusieurs valeurs de bande basse et de pleine bande et on pourra ainsi comparer les résultats.

En ce qui concerne les améliorations à apporter à nos algorithmes, nous avons donné quelques pistes dans le chapitre 5. De plus le système d'enrichissement par modèle de signal est incomplet, puisqu'il manque toute la partie de l'enrichissement des sinusoïdes.

Bibliographie

- [1] B. Bessette, R. Salami, R. Lefebvre, M. Jenilek, J. Rotola-Pikkila, J. Venio, H. Mikkola, and K. Järvinen. The adaptive multirate wideband speech codec (amr-wb). *IEEE Transactions on Speech and Audio Processing*, 10(8) :620–636, 2002.
- [2] M. Dietz, L. Liljeryd, K. Kjörling, and O. Kunz. Spectral band replication, a novel approach in audio coding. *112th AES Convention*, May 2002.
- [3] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000.
- [4] S. Essid, G. Richard, and B. David. Efficient musical instrument recognition on solo performance music using basic features. In *AES 25TH International conference, London, United Kingdom*, 2004.
- [5] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka. Rwc music database : Popular, classical, and jazz music databases. In *In Proc. 3rd International Conference on Music Information Retrieval*, pages 287–288, 2002.
- [6] R. Hubert and B. Kollmeier. Pemo-q 8212 : A new method for objective audio quality assessment using a model of auditory perception. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(6), 2006.
- [7] E. Larsen, R.M. Aarts, and M. Danessis. Efficient high-frequency bandwidth extension of music and speech. *Proceedings of the 112th AES Convention*, 2002.
- [8] D.D. Lee and H.S. Seung. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing 13 (Proc. NIPS*2000)*, 2001.
- [9] G. Miet. *Towards Wideband Speech by Narrowband speech bandwidth extension : magic effect or wideband recovery ?* PhD thesis, Laboratoire d'Acoustique de l'Université du Maine et Philips Consumer Communications, Le Mans, France, November 2001.
- [10] M. Nilsson, H. Gustafsson, S.V. Andersen, and W.B. Kleijn. Gaussian mixture model based mutual information estimation between frequency

- bands in speech. *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on*, 1 :525–528, 2002.
- [11] S. Oh, W. Yoon, Y. Cho, K. Park, and K. Kim. A new spectral enhancement algorithm in mp3 audio. *IEEE Transactions on Consumer Electronics*, 52(1) :196–199, 2006.
- [12] T. Virtanen. Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3), 2007.
- [13] X. Zhu, G. Beauregard, and L. Wyse. Real-time signal estimation from modified short-time fourier transform magnitude spectra. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(5), 2007.