Master 2 Sciences de l'ingénieur
Université Pierre et Marie Curie - Paris 6

parcours ATIAM

Rapport de stage de fin d'étude

Transcription automatique de musique polyphonique pour piano solo par factorisation du spectrogramme en matrices non-négatives

---

Auteur :

Antoine Falaize-Skrzek

Responsables de stage :

Bertrand David - Télécom Paristech
Laurent Daudet - ESPCI Paristech
François Rigaud - Télécom Paristech

23 août 2012

Le présent document rapporte les résultats obtenus pendant mon stage de fin d'études. Ceci constitue une étape importante dans le parcours d'un individu, et je n'aurais pu espérer meilleure structure que Télécom Paristech pour achever un cursus finalement dirigé vers les méthodes informatiques pour le traitement du signal. Je tiens donc particulièrement à remercier Bertrand David, tout d'abord pour sa confiance dans ma capacité à mener à bien cette étude, me permettant ainsi d'intégrer le groupe AudioSig de Télécom Paristech, mais surtout pour m'avoir fait bénéficier de son expérience pointue dans les domaines abordés (les signaux musicaux inspirés par la physique, les représentations temps/fréquence et l'analyse automatique).

Ce travail s'intègre dans celui de François Rigaud, actuellement en thèse de doctorat au sein du service traitement du signal et de l'image de Télécom Paristech, que je tiens à remercier vivement pour avoir avoir partagé avec moi sa longue réflexion sur les méthodes abordées, ainsi que pour la patience dont il a fait preuve pour répondre à mes nombreuses interrogations.

Je n'ai eu l'occasion qu'à deux reprises de renconter Laurent Daudet (ESPCI), que je remercie pour le profond intérêt qu'il a porté à mon travail, ainsi que pour ses remarques pertinentes sur le système proposé, nourries par son implication de longue date dans des projets similaires.

Par ailleurs, ce travail n'aurait aucune valeur si les résultats n'étaient pas comparés à des méthodes robustes. Ainsi, je remercie Benoit Fuentes et Valentin Émiya pour m'avoir fourni leurs codes de transcription automatique.

Les quelques échanges que j'ai pu avoir avec Roland Badeau et Gaël Richard ont beaucoup contribué à faire avancer ma propre réflexion. Aussi les remercié-je pour le temps qu'ils m'ont consacré.

Enfin, je tiens à remercier chaleureusement toute l'équipe du groupe AudioSig (permanents, post-doctorants, doctorants et stagiaires), de l'accueil qui m'a été fait, et de l'aide apportée pour mes démarches fructueuses en vue d'obtenir une bourse de thèse. Cotoyer tant d'intelligence durant ces cinq mois de stage m'a apporté un autre regard sur le traitement de signal, la recherche scientifique - et les rapports humains en règle générale.

Antoine Falaize-Skrzek

Table des matières

Notations

Mathématiques

$\mathcal{M}_{m,n}(\mathbb{R})$	Ensemble des matrices de m lignes et n colonnes à coefficient dans $\mathbb{R}$
$d_\beta(.\|.)$	β-divergences
$\mathbf{V}$	Matrice
V_m	Vecteur ligne m de $\mathbf{V}$
V_{mn}	Élément ligne m, colonne n de $\mathbf{V}$

Algorithmes

$\mathbf{V}$	Spectrogramme
$\mathbf{W}$	Dictionnaire (approximation *NMF)*
$\mathbf{H}$	Activations (approximation *NMF)*
`Ha`	Paramétrisation harmonique de $\mathbf{W}$
`Ih`	Paramétrisation inharmonique de $\mathbf{W}$
`IhR`	Contrainte relaxée d'inharmonicitée sur $\mathbf{W}$

Abréviations

TFCT	Transformée de Fourier à court terme
NMF	Non-negative Matrix Factorization
AR	Modèle Auto-Régressif
MA	Modèle à moyenne ajustée

Introduction

Le présent document rapporte le travail effectué durant mon stage de fin de Master 2 Sciences De l'Ingénieur parcours *Acoustique et Traitement Informatique Appliqué à la Musique* de l'*Université Pierre et Marie Curie*, en partenariat avec l'IRCAM. Ce stage s'est déroulé dans le département *Traitement du Signal et de l'Image*, l'un des quatre départements d'enseignement et de recherche de l'école *Télécom Paristech*. Dans sa composante recherche, il fait partie de l'UMR CNRS 5141 LTCI. Les principaux thèmes de recherche sont : le développement d'algorithmes et de traitements statistiques, en particulier pour l'apprentissage de modèles, l'indexation multimédia et le codage, ainsi que la transmission des différentes communications multimédias.

Le sujet de ce stage est l'évaluation des performances de transcription de différents algorithmes basés sur la factorisation du spectrogramme en matrices non-négatives. Le système développé est représenté figure 1. Ce travail se rattache à la thèse de François Rigaud, qui s'entreprend à développer une variante contrainte de la méthode *NMF* (pour *Non-negative Matrix Factorization*) permettant l'estimation des paramètres d'inharmonicité sur l'ensemble de la tessiture du piano. Ainsi, l'objectif du présent document est de répondre à la question suivante :

Quel est l'apport des modèles d'inharmonicité pour la tâche de transcription aveugle de musique polyphonique pour piano solo ?

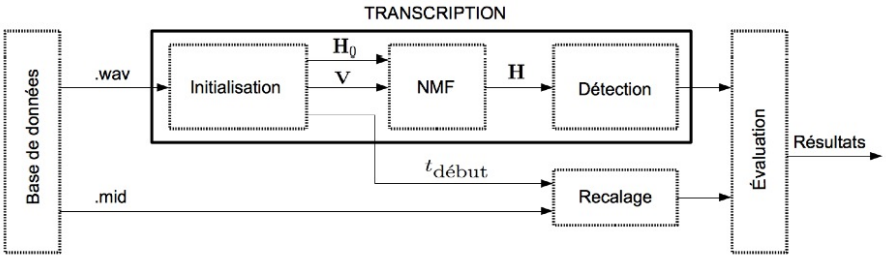
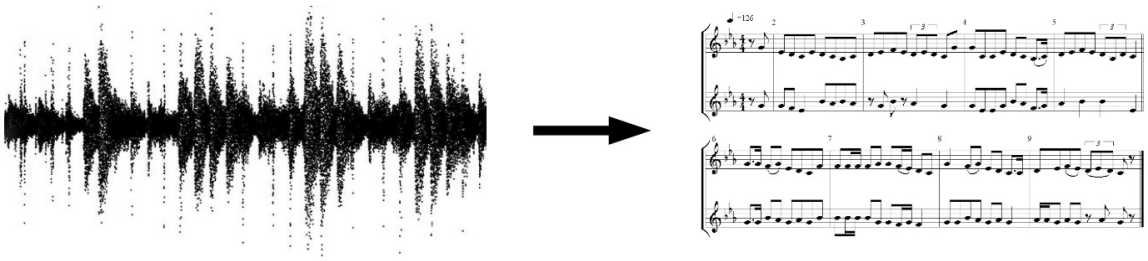


Figure 1 – Schématisation du système développé dans son ensemble.

La transcription automatique tient une place importante dans le domaine de la recherche d'informations musicales. Pour définition, on cite ici Nancy Bertin, dans [Bertin, 2009] :

Par le mot transcription, on entend généralement une opération de conversion, de transformation d'une représentation de certaines données d'un domaine vers un autre, et en particulier, d'un domaine de bas-niveau, peu descriptif, vers un domaine de haut niveau, sémantiquement plus riche.



Ainsi, la tâche consiste à passer une oeuvre musicale de sa représentation en *forme d'onde* échantillonnée à une description de plus haut niveau incluant des informations sur chaque note, telles le timbre, la hauteur, la durée, ce que l'on retrouve dans la *partition de musique* usuelle. Le pionier dans le domaine est sans doute James Andy Moorer qui posa les bases de cette discipline à la fin des années 70 (*cf.* [Moorer, 1977]). Les principales restrictions posées à l'époque afin de résoudre le problème sont la limitation de la polyphonie à deux voix et un rapport inharmonique entre deux notes simultanées.

Valentin Emiya donne dans [Emiya, 2008] une description détaillée des méthodes proposées à la suite des travaux de J.A. Moorer. Les systèmes actuels reposent essentiellement sur deux approches :

- **l'estimation de hauteur,** sous l'hypothèse que l'information liée à la hauteur des notes présentes peut être extraite d'une description fréquentielle instantanée (de l'ordre d'une trame) ;

- **les modèles de notes,** qui s'attachent à décrire les observations directement en des termes de critères musicaux (approches bayésiennes, dictionnaires de formes d'ondes). Les dimensions temporelle et fréquentielle sont alors utilisées conjointement.

La transcription par méthode *NMF* permet de traiter le problème selon l'une ou l'autre des approches, puisque l'on vient estimer les paramètres de modèles (déterministes ou probabilistes), posés pour chaque point temps/fréquence. L'attention est portée sur la musique pour piano, car cet instrument présente un haut degré de polyphonie (de l'ordre de la dizaine de notes) sur une grande tessiture (88 notes sur un piano standard). De plus, les sons de piano sont fortement inharmoniques, ce qui en fait un défi particulièrement intéressant. Nous nous proposons d'inclure une connaissance sur la physique du piano afin de réaliser la transcription, ce qui inscrit ce travail dans la continuité des approches proposées dans [Emiya, 2008] et [Durrieu, 2010].

Les méthodes *NMF* ont fait l'objet d'une littérature abondante, et sont tout d'abord présentées §1, avec les bases de la physique des *sons de piano*. Nous détaillons ensuite §2 les variantes contraintes de la *NMF* utilisées en coeur du système de transcription puis l'ensemble *pré/post*-traitement qui fait l'objet du chapître §3. La base de donnée, la procédure d'évaluation et les résultats obtenus sont donnés chapitre §4, avant la conclusion.

Chapitre 1

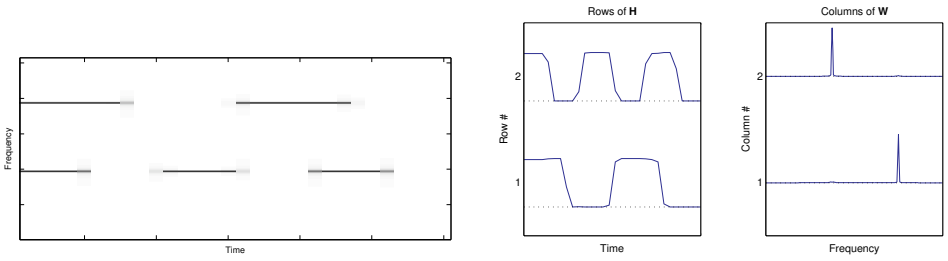
État de l'art

Nous présentons ici un état de l'art de la méthode *NMF*, puis nous précisons les spécificités des *sons de piano*.

1.1 La factorisation en matrices non-négatives

Les bases de cette méthode d'estimation de paramètres à la vue d'observations ont été posées en 1994 dans [Paatero and Tapper, 1994], sous l'appellation "*Positive Matrix Factorization*". Le terme *NMF* (pour "*Non-negative Matrix Factorization*") apparaît en 1999 dans l'article [Lee and Seung, 1999], cité aujourd'hui en référence, dans lequel les auteurs proposent d'appliquer la méthode à la décomposition d'images, en précisant qu'il est possible de traiter des problèmes d'une grande variété de domaines. Il est proposé de décomposer une matrice positive $\mathbf{V} \in \mathcal{M}_{m,n}(\mathbb{R}_+)$ en produit de deux matrice $\mathbf{W} \in \mathcal{M}_{m,r}(\mathbb{R}_+)$ et $\mathbf{H} \in \mathcal{M}_{r,n}(\mathbb{R}_+)$. On notera que le cas $mr+rn < mn$ traduit une réduction de la dimension de $\mathbf{V}$ à la condition que l'erreur de reconstruction soit nulle (*cf.* [Bertin, 2009]).

La *NMF* est décrite dans [Lee and Seung, 1999] comme une méthode de modélisation de processus génératifs de variables directement observables $\mathbf{V}$, à partir de variables cachées $\mathbf{H}$. Chacune de ces variables cachées coactive un sous ensemble de variables visibles (atomes) contenues dans $\mathbf{W}$ (qui s'apparente donc à un dictionnaire). L'activation de plusieurs variables cachées combine les atomes additivement pour créer un tout approchant les observations.



(a) Spectrogramme initial $\mathbf{V}$

(b) Matrice d'activation $\mathbf{H}$ et dictionnaire $\mathbf{W}$ (de gauche à doite)

FIGURE 1.1 – Exemple simple d'estimation pour $r = 2$ (tiré de [Smaragdis, 2003])

On trouvera dans [Smaragdis, 2003] la première application de la *NMF* à la transcrip-

tion automatique de musique polyphonique. L'objectif est d'approcher le spectrogramme V_{ft} issu du module au carré de la TFCT par la superposition de spectres de base dont l'amplitude varie dans le temps $\tilde{V}_{ft}$ (*cf.* figure 1.1).

$$V_{ft} \approx \tilde{V}_{ft} = \sum_{r=1}^{R} W_{fr} H_{rt}; \quad t \in [1, T], \quad f \in [1, F], \tag{1.1}$$

avec F le nombre de canaux de la TFCT, T le nombre de trames du spectrogramme et R le nombre d'atomes du dictionnaire $\mathbf{W}$ (qui pilote la profondeur de la décomposition). La matrice $\mathbf{H}$ des activations de chaque *note* dans le temps est alors proche d'un *piano-roll midi*, duquel il est possible de tirer une partition (via des systèmes dédiés, qui ne sont pas étudiés ici).

1.1.1 Fonction de coût

La factorisation de type *NMF* est obtenue par la minimisation d'une fonction de coût $\mathcal{C}$, définie comme une distance matricielle :

$$\mathcal{C}(\mathbf{V}|\mathbf{WH}) = D(\mathbf{V}|\tilde{\mathbf{V}}). \tag{1.2}$$

En considérant chaque point du spectrogramme indépendant des autres, la métrique posée pour le calcul de la distance $D(.|.)$ est la somme

$$\mathcal{C}(\mathbf{V}|\mathbf{WH}) = \sum_{f=1}^{F} \sum_{t=1}^{T} d(V_{ft}|\tilde{V}_{ft}), \tag{1.3}$$

avec $d(a|b)$ une fonction de deux variables scalaires, à valeurs dans $\mathbb{R}_+$, qui s'annule pour $a = b$. Une propriété intéressante pour la fonction de coût est la convexité (soit le fait que la fonction possède un minimum global). Ainsi, il sera possible de prouver la décroissance du coût de l'approximation au fur et à mesure des itérations [1].

Les distances/divergences utilisées dans la littérature sont :

- la distance euclidienne $d_{EUC}(a|b) = \frac{1}{2}(a-b)^2$
- la divergence généralisée de Kullback-Lieber $d_{KL}(a|b) = a\log(\frac{a}{b}) - a + b$
- la divergence d'Itakura-Saito $d_{IS}(a|b) = \frac{a}{b} - \log(\frac{a}{b}) - 1$

De nombreux travaux mentionnent la β-divergence, définie comme (*cf.* [Bertin, 2009]) :

$$d_\beta(a|b) = \begin{cases} \frac{1}{\beta(\beta-1)}(a^\beta + (\beta-1)b^\beta - \beta a b^{\beta-1}) & \beta \in \mathbb{R}\backslash\{0,1\}, \\ a\log(\frac{a}{b}) + (b-a) & \beta = 1, \\ \frac{a}{b} - \log(\frac{a}{b}) - 1 & \beta = 0. \end{cases} \tag{1.4}$$

Cependant, on trouve un argument pour utiliser une autre fonction de coût. En effet, les signaux musicaux présentent une forte dynamique (de l'ordre de 100dB). Ainsi le *résiduel relatif*

$$R_{ft} = \frac{V_{ft} - [WH]_{ft}}{V_{ft}} \tag{1.5}$$

1. Cependant, plusieurs auteurs (par exemple N. Bertin dans [Bertin, 2009] §IV.2.1.2) font remarquer qu'une valeur de coût plus faible ne correspond pas nécessairement à une sémantique de la représentation plus pertinente (c'est à dire aisément interprétable, ou sur la base de laquelle l'estimation des fréquences fondamentales et des *onsets* est simplifiée).

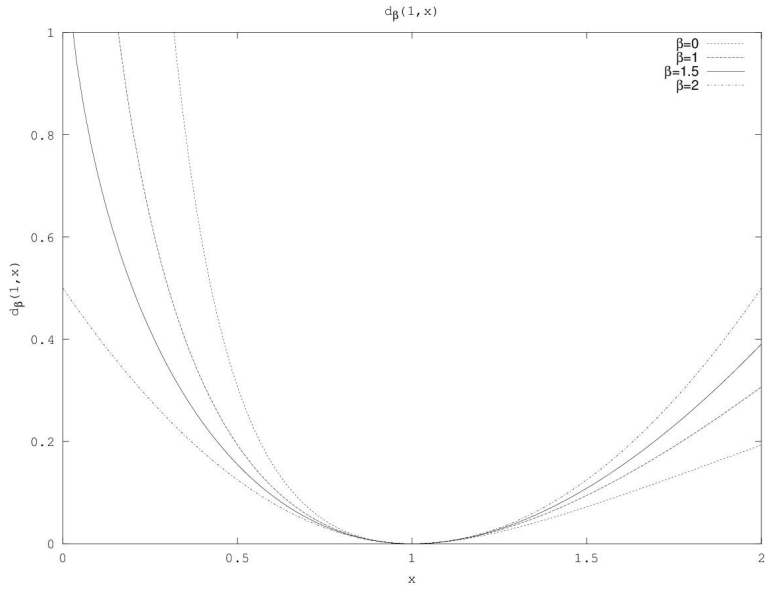


FIGURE 1.2 – Allure des β-divergences ($d_\beta(1|x)$ équation 1.4) pour différentes valeurs de β

peut devenir acceptable pour les notes de grande énergie, et pénaliser la mise à jour pour les notes plus faibles. Pour remédier à cela, E. Vincent propose dans [Vincent et al., 2007] §2.2 l'utilisation de la norme euclidienne pondérée, qui intègre une pondération tenant compte des attributs perceptifs (les règles de mise à jour sont données dans [Virtanen, 2004] §3) et précise dans [Vincent et al., 2009] §2.2 qu'une valeur faible de β se traduit par une compression de la dynamique du spectre, améliorant de fait les performances de modélisation des parties de faible amplitude.

1.1.2 Type d'algorithmes

On trouvera dans [Cichocki et al., 2009] un inventaire des différents algorithmes utilisables. Cependant, dans la mesure où les algorithmes multiplicatifs garantissent la positivité des matrices **W** et **H** au fil des itérations, l'attention est profondément portée sur ces derniers [2]. [Hennequin et al., 2010] §3 donne un aperçu de la manière d'obtenir les règles de mise à jour d'un paramètre ψ (tous les autres sont donc fixés) pour une fonction de coût $\mathcal{C}$ dans le cadre d'un algorithme multiplicatif de descente. La dérivée partielle de $\mathcal{C}$ par rapport à ψ est décomposée en la différence de deux termes positifs :

$$\frac{\partial \mathcal{C}}{\partial \psi} = \Psi_+ - \Psi_- \tag{1.6}$$

ce qui amène à la règle de mise à jour 1.7, qui assure que l'évolution du paramètre se fait dans le sens du gradient, et que le paramètre reste positif :

$$\psi \leftarrow \psi \frac{\Psi_-}{\Psi_+} \tag{1.7}$$

2. Cependant, [Vincent et al., 2009] §2.2 fait remarquer que la preuve de la décroissance des β-divergences sous les règles multiplicatives de mise à jour des matrices **W** et **H** n'est donnée que pour $1 \leq \beta \leq 2$, mais qu'en pratique la convergence est constatée $\forall \beta$.

On notera que la dérivée de $d_\beta(a(\psi_1)|b(\psi_2))$ (*cf.* [1.4]) par rapport à l'une ou l'autre des variables s'écrit :

$$\begin{aligned} \frac{\partial d_\beta}{\partial \psi_1} &= \frac{\partial a}{\partial \psi_1} \frac{a^{\beta-1} - b^{\beta-1}}{\beta - 1} \\ \frac{\partial d_\beta}{\partial \psi_2} &= \frac{\partial b}{\partial \psi_2} b^{\beta-2}(b-a) \end{aligned} \tag{1.8}$$

L'ensemble des approches visant à contraindre la NMF prend comme départ l'ajout d'un terme de pénalité $\mathcal{C}_\lambda$ au calcul du coût de la factorisation. Ce que l'on résume ainsi (*cf.* [Bertin, 2009]) :

$$\min_{\mathbf{W},\mathbf{H}} \mathcal{C} = \min_{\mathbf{W},\mathbf{H}} \left[\mathcal{C}_0(\mathbf{V}|\mathbf{WH}) + \lambda \mathcal{C}_\lambda(\mathbf{W},\mathbf{H})\right] \tag{1.9}$$

où λ est un terme de pondération, c'est à dire le poids relatif de la pénalité face au coût standard de la factorisation (croît en même temps que le produit $\mathbf{WH}$ se rapproche de V).

Parmi les contraintes les plus utilisées, on citera la parcimonie (de $\mathbf{W}$ et de $\mathbf{H}$), la régularité (des lignes de $\mathbf{H}$), et la décorrélation (des colonnes de $\mathbf{W}$).

Une approche consistant à paramétrer directement la forme du dictionnaire est envisagée dans [Vincent et al., 2009] et [Hennequin et al., 2010]. Ce sont alors les paramètres qui sont mis à jour par NMF, et non le dictionnaire directement. La méthode développée s'appuie sur celle décrite dans [Hennequin et al., 2010], détaillée ci-après.

1.1.3 Dictionnaire W harmonique

La paramétrisation du dictionnaire est réalisée en posant un modèle sous chaque colonne de $\mathbf{W}$. On doit cet algorithme à R. Hennequin (*cf.* [Hennequin et al., 2010]), qui propose une évolution temporelle des atomes qui n'est pas reprise ici (puisque le piano ne permet pas de modifier instantanément la fréquence des notes, ce que permet la guitare, un autre instrument à cordes en oscillations libres, par les *bends*). Il est proposé d'approcher le module au carré de la TFD d'ordre N dans le canal f à la trame t (V_{ft}) par $\tilde{V}_{ft}$, tel que

$$V_{ft} \approx \tilde{V}_{ft} = \sum_{r=1}^{R} W_{fr}^{\theta_r} H_{rt}, \tag{1.10}$$

où θ_r est le paramètre associé à l'atome r. Le paramètre est ici la fréquence fondamentale de chaque atome harmonique. On choisit de définir le dictionnaire sur la base de la gamme chromatique : $\theta_r = F_{0,r} = 2^{\frac{r-1}{12}} F_{\text{ref}}$, avec F_{ref} la fréquence la plus basse du dictionnaire $\mathbf{W}$ (ainsi, pour couvrir toute la tessiture du piano, on choisira $F_{\text{ref}} = 27.5Hz$ et $R = 88$) :

$$W_{fr}^{\theta_r} = \sum_{k=1}^{K_r} a_k \, g(f - k\, F_{0,r}), \tag{1.11}$$

ce qui correspond au module du spectre d'une somme pondérée de K_r sinusoïdes fenêtrées, élevé au carré. On notera que les interférences entre partiels sont négligées (hypothèse que f_0^{rt} est suffisamment élevée, ou, de manière équivalente, que la taille de la fenêtre utilisée dans la TFct est suffisamment grande). K_r est défini comme le nombre maximum de partiels de l'atome r qui assure qu'ils sont tous sous une fréquence maximum choisie au préalable (généralement $F_{max} = \frac{F_e}{2}$) :

$$K_r = \{\max(k) \mid 2^{\frac{r-1}{12}}\, k\, F_{ref} < F_{max}\} \tag{1.12}$$

g est le module au carré de la transformée de Fourier de la fenêtre d'analyse utilisée pour calculer le spectrogramme. Pour une fenêtre de *Hann* de durée $\mathcal{T}$ son expression est :

$$g(f) = \left(\frac{\sin(\pi \mathcal{T} f)}{2\pi f(\mathcal{T}^2 f^2 - 1)} \right)^2 . \tag{1.13}$$

Les amplitudes a_k de chaque partiel sont supposées identiques pour chaque atome (afin de s'affranchir d'éventuelles erreurs d'octave induites par la mise à zéro d'un partiel sur deux). On trouvera le détail des règles de mise à jour dans B.1. L'algorithme proposé dans [Hennequin et al., 2010] utilise en parallèle une version standard de la NMF, afin de prendre en compte les composantes bruitées du signal initial, évincées par définition des atomes paramétriques harmoniques. Cette branche de la méthode n'est pas implémentée ici. On fera référence à la paramétrisation harmonique de $\mathbf{W}$ sous l'appellation `Ha`. Un exemple de résultats obtenus en terme de mise à jour des paramètres F_0 et a_k est donné figure 1.3.

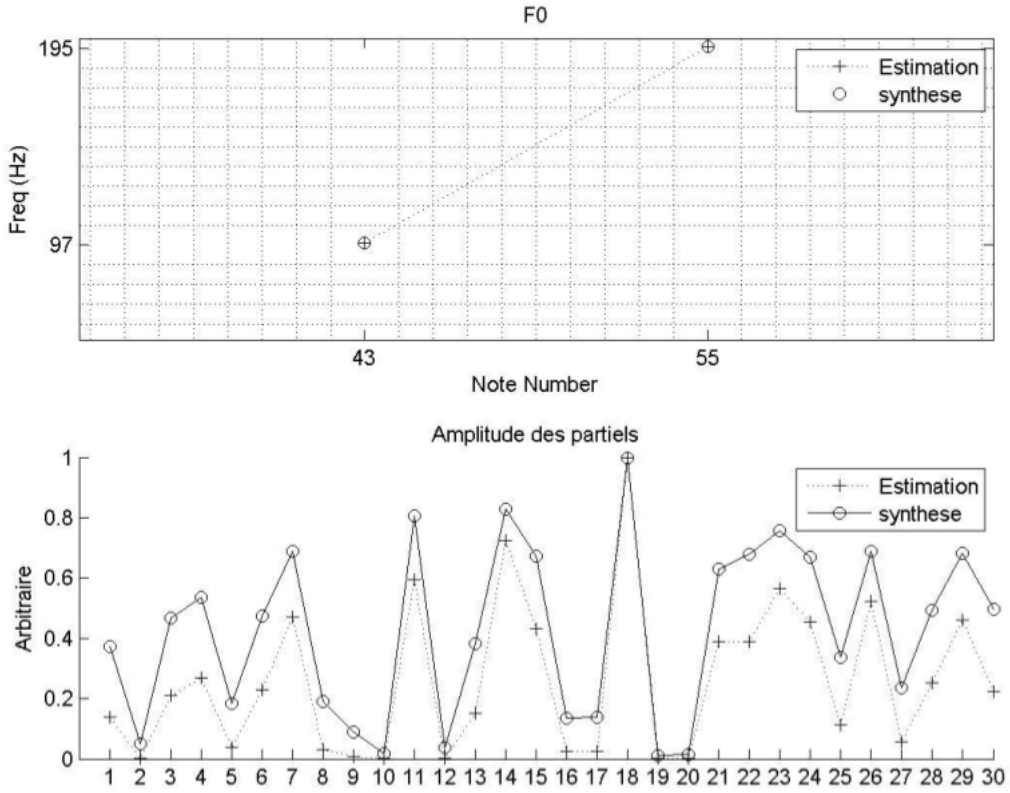
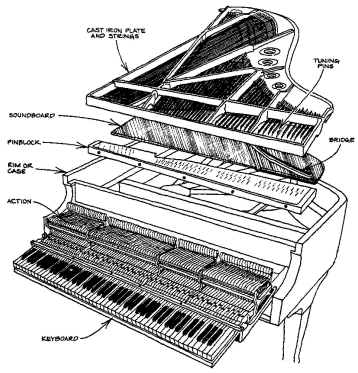


FIGURE 1.3 – Fréquences fondamentales et amplitudes des partiels synthétisées et estimées par la méthode `Ha` pour deux notes de piano (Midi note numbers 43 et 55, générés avec 30 partiels). On constate que les fréquences fondamentales sont correctement estimées (figure du haut), et que les amplitudes des partiels sont qualitativement proches de la synthèse (figure du bas).

Nous présentons section suivante les spécifités des *sons de piano* qui nous ont poussé à introduire une paramétrisation inharmonique de $\mathbf{W}$.

1.2 Modélisation de la note de piano

Le piano moderne est une évolution du *piano-forte* de Bartolomeo Cristofori, inspiré du clavicorde. Tous ces instruments appartiennent à la famille des instruments à cordes frappées en oscillations libres. La production du son pour le piano peut être succintement décrite comme suit (*cf.* [Emiya, 2008], [Fletcher and Rossing, 1998]) :



- la corde initialement au repos subit l'impact du marteau, qui donne l'excitation de départ.
- après une série de phénomènes complexes de premières réflexions de l'onde de choc sur les extrémités, la corde entre en oscillation libre
- la vibration est transmise par couplage mécanique de la corde à la table d'harmonie
- la table d'harmonie met en vibration le milieu ambiant par rayonnement (couplage mécano-acoustique)
- enfin, le milieu ambiant transmet par conduction aérienne la perturbation jusqu'au récepteur (oreilles de l'auditeur ou microphone).

On notera que deux à trois cordes légèrement désaccordées sont excitées pour chaque note jouée, et que la vibration transversale se fait suivant les deux polarisations possibles. Ainsi chaque partiel d'une note peut combiner jusqu'à six pics de fréquences très proches, et les différents couplages induisent une évolution temporelle individuelle de l'amplitude de chaque partiel (*cf.* figure 1.4).

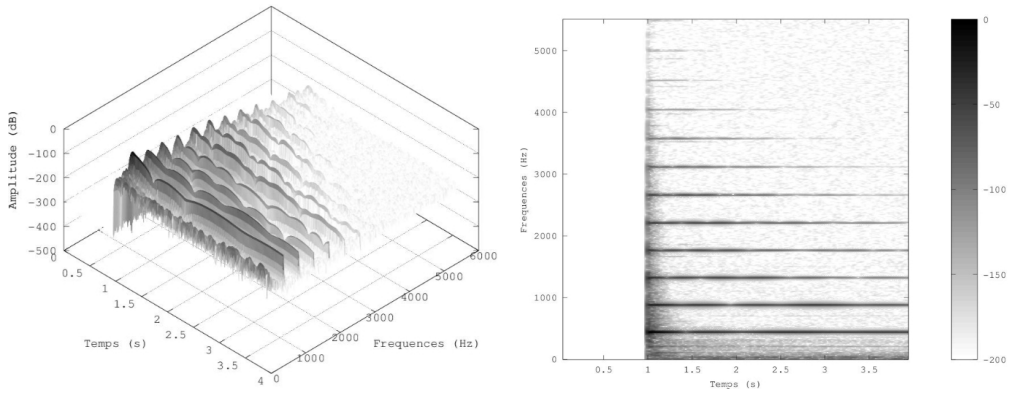
L'impact du marteau et la mise en oscillation des cordes se traduisent par un bruit large bande qui se dissipe rapidement (sur un temps de l'ordre de la dizaine de millisecondes). On constate que l'atténuation de l'amplitude de chaque partiel (propre aux instruments à excitation non-entretenue) se fait suivant une pente exponentielle, dont le coefficient de décroissance est indépendant des autres partiels. Cela rend difficile une description complète de l'évolution de chaque partiel, de chaque note, de chaque piano. Aussi nous interesserons-nous au modèle donné dans [Fletcher and Rossing, 1998].

En première approximation, la fréquence de chaque partiel est indépendante du couplage avec la table d'harmonie. Ainsi, N. H. Fletcher et T. D. Rossing proposent la solution de l'équation de la corde avec raideur écrite en deux dimensions (équation 1.14) comme expression analytique des fréquences recherchées :

$$\rho \frac{\partial^2 y}{\partial t^2} = T \frac{\partial^2 y}{\partial x^2} - EI \frac{\partial^4 y}{\partial x^4} \tag{1.14}$$

avec y le déplacement transversal de la corde (en mètres), x la position dans la direction longitudinale (en mètres), ρ la densité linéique de masse (en $kg.m^{-1}$), T la tension (en Newtons), E le modle d'Young du matériau dans lequel la corde est fabriquée, $I = \frac{\pi d^4}{64}$ le moment quadratique d'une section circulaire (avec d le diamètre de la corde en mètres) et t le temps (en secondes). Sous l'hypothèse que les extrémités de la corde sont fixes (on néglige le couplage avec la table d'harmonie), les fréquences de chaque mode k solution de [1.14] sont données par

$$f_k = kF_0\sqrt{1 + Bk^2} \tag{1.15}$$



(a) Vue de côté (b) Vue de dessus

FIGURE 1.4 – Spectrogramme normalisé et exprimé en décibels des trois premières secondes d'une note de piano (La_4, $440Hz$, d'un piano DISKLAVIER tiré de la base *MAPS*) échantillonée à $44100Hz$, calculée avec une fenêtre de Hann de 2^{12} points et un incrément de 512 points sur les trames d'analyse

avec :

- $F_0 = \frac{1}{2L}\sqrt{\frac{T}{\rho}}$ la fréquence de la corde sans raideur associée (en Hertz),
- L la longeur de la corde (en mètres)
- $B = \frac{\pi^3 E d^4}{64TL}$ le coefficient d'inharmonicité
- $k \in \mathbb{N}^*$ l'orde du partiel

Pour la suite, seule la distribution des fréquences des partiels donnée par l'équation (1.15) sera utilisée. Nous présentons au chapitre suivant les variantes contraintes de la *NMF* utilisées en coeur du système de transcription proposé.

Chapitre 2

Méthode développée

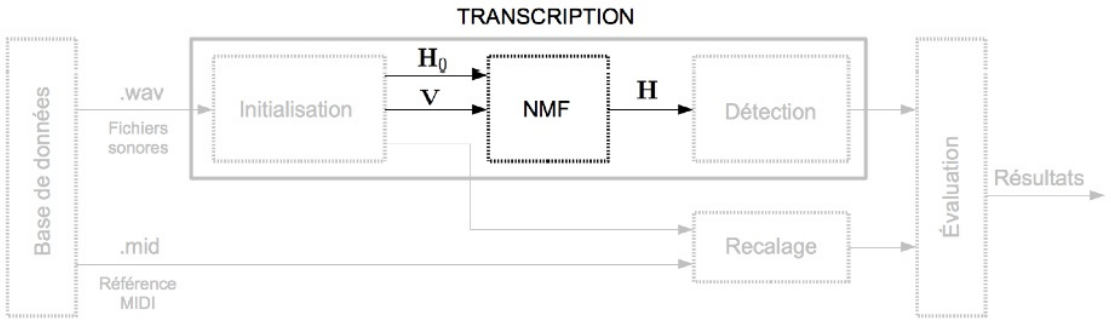


FIGURE 2.1 – Schématisation du système de transcription

Comme expliqué §1.1.3 et §1.2, nous utilisons pour notre système une paramétrisation inharmonique du dictionnaire **W**. François Rigaud propose dans [Rigaud et al., 2011] deux paramétrisations construites à partir de celle de R. Hennequin (donnée §1.1.3), en incluant le modèle de distribution inharmonique des fréquences des partiels donné par l'équation 1.15. Ces paramétrisations font l'objet de la section 2.1, à la différence que nous travaillons sur le spectrogramme (module au carré de la TFCT), ce qui modifie les règles de mise à jour (les nôtres sont données en annexes B.2 et B.3).

Étant donné que nous nous concentrons sur la matrice d'activation pour la transcription, nous décrivons aussi §2.2 un ensemble de contraintes sur **H** mises en place afin d'améliorer les performances.

2.1 Dictionnaire W inharmonique

Une première paramétrisation pour laquelle la loi d'inharmonicité (équation 1.12) est inclue directement dans la forme des atomes équation 1.11 est décrite. Ensuite, elle est exprimée sous forme de contrainte relaxée.

2.1.1 Contrainte stricte d'inharmonicité

θ_r est ici la fréquence de référence et le coefficient d'inharmonicité de chaque atome inharmonique $\theta_r \sim [F_{0,r}, B_r]$. La loi d'inharmonicité que suit le partiel de rang k de chaque

atome est donnée par l'équation 1.15, d'où la paramétrisation du dictionnaire **W** :

$$W_{fr}^{\theta_r} = \sum_{k=1}^{K_r} a_k \, g(f - f_{r,k}), \tag{2.1}$$

avec $f_{r,k}$ la fréquence du *k-ième* partiel de l'atome r. L'expression du nombre maximum de partiels $K_r \mid f_{K_r} < F_{\max}$ est alors (dans le cas où le dictionnaire suit la loi d'inharmonicité de coefficients moyens $[\hat{F}_{0,r}, \hat{B}_r]$) :

$$K_r(\theta_r) = \left[\frac{\sqrt{1 + 4B_r \frac{F_{\max}}{F_{0,r}}}}{2B_r} \right]^{1/2} \tag{2.2}$$

L'ensemble des règles de mise à jour est donné dans B.2. On fera référence à la paramétrisation inharmonique de **W** sous l'appellation `Ih`. Un exemple de résultat est donné figure 2.2. On constate que les paramètres ne sont pas correctement mis à jour, ce qui est la raison de l'inclusion de la contrainte d'inharmonicité sous forme d'un terme supplémentaire dans la fonction de coût, §2.1.2.

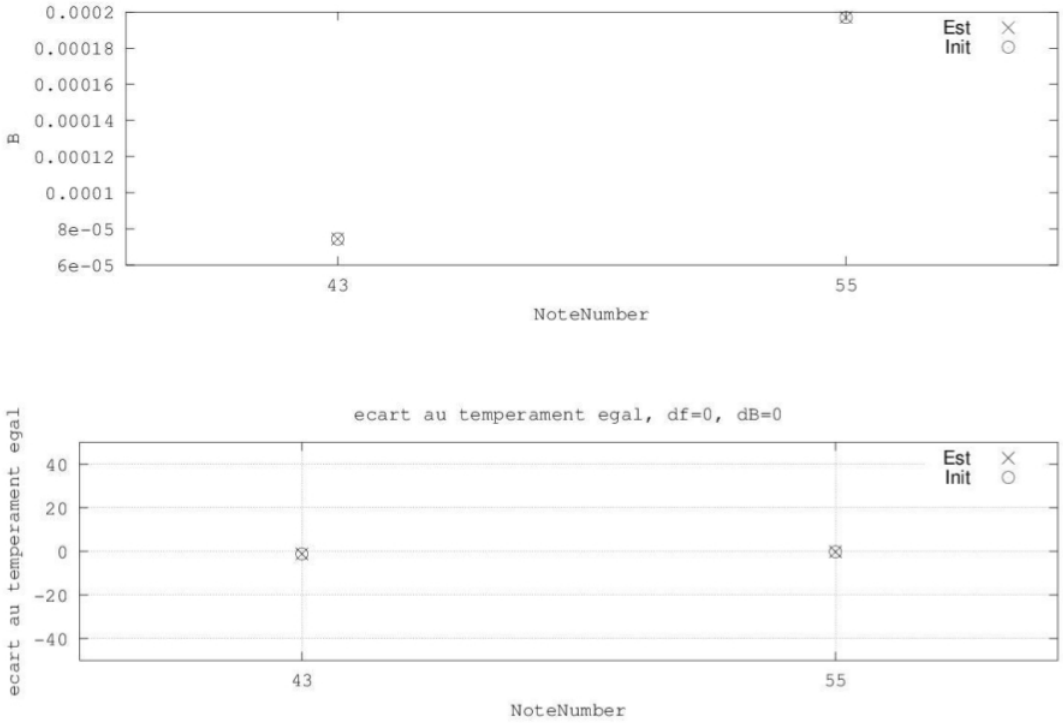


FIGURE 2.2 – Coefficients d'inharmonicité (B_r, en haut) et fréquences de référence ($F_{0,r}$, en bas, présentées en écart au tempérament égal exprimés en cents) estimés par la méthode `Ih` pour deux notes de piano (Midi note numbers 43 et 55, tirées de la base *MAPS*). On constate que les paramètres restent "bloqués" sur les valeurs de l'initialisation.

2.1.2 Contrainte relaxée d'inharmonicité

θ_r est ici l'ensemble des fréquences et des amplitudes des K_r partiels de chaque atome, pouvant s'adapter indépendamment aux données $\theta_r \sim [a_{r,k}, f_{r,k}], \ \forall k \in [1, K_r]$. La pa-

ramétrisation du dictionnaire $\mathbf{W}$ est donc

$$W_{fr}^{\theta_r} = \sum_{k=1}^{K_r} a_{r,k}\, g(f - f_{r,k}). \tag{2.3}$$

Pour que les partiels $f_{r,k}$ tendent à suivre la loi inharmonique, on introduit dans la fonction de coût un terme supplémentaire :

$$\begin{aligned} \mathcal{C}(\Theta, \gamma, H) &= \mathcal{C}_0(\Theta, H) + \lambda_1 \mathcal{C}_1(f_{r,k}, \gamma) \\ \mathcal{C}_0(\Theta, H) &= \frac{1}{FT} \sum_{ft} d_{\beta_0}(V_{ft} | \tilde{V}_{ft}) \\ \mathcal{C}_1(f_{r,k}, \gamma) &= \sum_r \frac{1}{K_r} \sum_{k=1}^{K_r} d_{\beta_1}(f_{r,k} | k F_{0,r} \sqrt{1 + k^2 B_r}) \end{aligned} \tag{2.4}$$

où
– $\Theta = \{a_{r,k}, f_{r,k}\},\ r \in [1, R],\ k \in [1, K_r]$
– $H = \{H_{rt}\},\ r \in [1, R],\ t \in [1, T]$
– $\gamma = \{F_{0,r}, B_r\},\ r \in [1, R]$

On fera référence à la contrainte relaxée d'inharmonicité sur $\mathbf{W}$ sous l'appellation `IhR`, dont on trouvera les règles de mise à jour dans B.3. Un exemple d'estimation est donné figure 2.3. Cette méthode semble mieux s'adapter aux données. Les performances seront évaluées au chapitre §4.

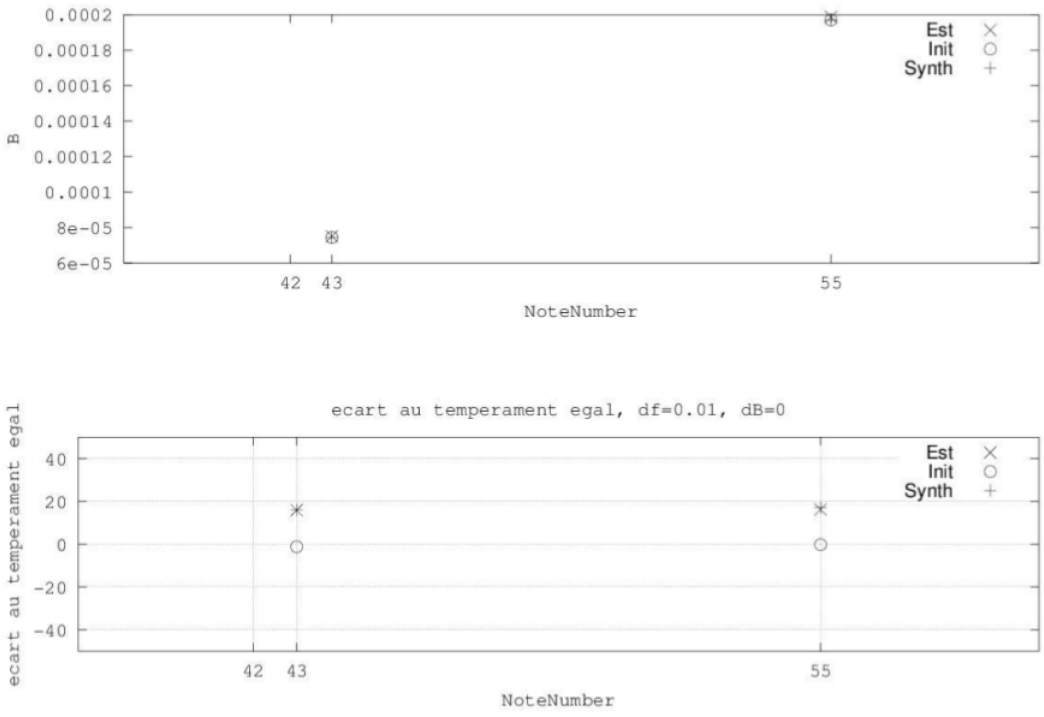


FIGURE 2.3 – Coefficients d'inharmonicité (B_r, en haut) et fréquences de référence ($F_{0,r}$, en bas, présentées en écart au tempérament égal exprimés en cents), à l'initialisation et estimés par la méthode `IhR` pour deux notes de piano (Midi note numbers 43 et 55, synthétisées). On constate que les paramètres sont correctement estimés.

La section suivante expose les différentes contraintes posées sur la matrice d'activation.

2.2 Contraintes sur H

Nous proposons ici d'inclure une mise en forme des lignes de **H** au fur et à mesure des itérations, afin de faciliter la détection, §3.2. Nous présentons tout d'abord deux pénalités visant à forcer la régularité des lignes et la parcimonie des colonnes de **H**, puis une paramétrisation des activations sous forme d'exponentielles décroissantes.

2.2.1 Critère de régularité

Plusieurs auteurs ont proposé des méthodes pour contraindre les lignes de **H** à être régulières. D'un commun accord, quelle que soit la méthode, il est remarqué que la régularité temporelle diminue la capacité de la *NMF* à rendre compte des attaques, très prononcées dans le cas des instruments à oscillations libres, et qui sont par définition des "irrégularités". T. Virtanen propose dans [Virtanen, 2007] une méthode simple, basée sur la pénalisation des écarts d'amplitude d'une trame à l'autre, pour l'ensemble des piste d'activation. Cela amène à une contrainte relaxée, qui s'écrit comme suit :

$$C_T(\mathbf{H}) = \sum_{r=1}^{R} \frac{1}{\sigma_r^2} \sum_{t=2}^{T} (h_{r,t} - h_{r,t-1})^2 \tag{2.5}$$

où σ_r^2 est défini comme une estimation de l'écart type, qui permet de s'affranchir d'un simple rapport d'échelle (*ie.* $\mathbf{W} \longleftarrow \mathbf{W} * \text{Cste}$ et $\mathbf{H} \longleftarrow \mathbf{H}/\text{Cste}$) et se calcule par

$$\sigma_r^2 = \sqrt{\frac{1}{T} \sum_{t=1}^{T} h_{r,t}^2}. \tag{2.6}$$

Alors, la fonction de coût globale est la somme des fonctions de côut décrites §1.1.3 et 2.1, et de $\lambda_T C_T(\mathbf{H})$, avec λ_T le poids de la pénalité liée à la régularité des lignes de **H**. On constate figure 2.4 que la régularité seule n'apporte pas de progès significatifs, aussi proposons-nous la contrainte de parcimonie de la section suivante.

2.2.2 Critère de parcimonie

Initialement proposée en traitement de l'image (*cf.* [Olshausen and Field, 1997]), la parcimonie est largement utilisée en séparation de sources et décomposition aveugles. L'objectif est de limiter l'activation simultanée de plusieurs atomes, en rendant les colonnes de **H** parcimonieuses. On trouvera dans [Virtanen, 2007] §C l'expression de la pénalité liée à la profusion d'activations simultanées dans le cadre de la décomposition par *NMF* :

$$C_S(\mathbf{H}) = \sum_{r=1}^{R} \sum_{t=1}^{T} f\left(\frac{h_{r,t}}{\sigma_r}\right) \tag{2.7}$$

L'auteur indique cependant que cette contrainte n'apporte pas de progrès significatif pour l'application visée (la séparation de source). P.O. Hoyer propose dans [Hoyer, 2004] une mesure de la parcimonie d'un vecteur, construite sur le rapport entre les normes $\mathcal{L}_1$ et $\mathcal{L}_2$ de ce vecteur :

$$\text{parcimonie}(\vec{x}) = \frac{1}{\sqrt{n}-1} \left[\sqrt{n} - \frac{\sum_i |x_i|}{\sqrt{\sum_i x_i^2}} \right] \tag{2.8}$$

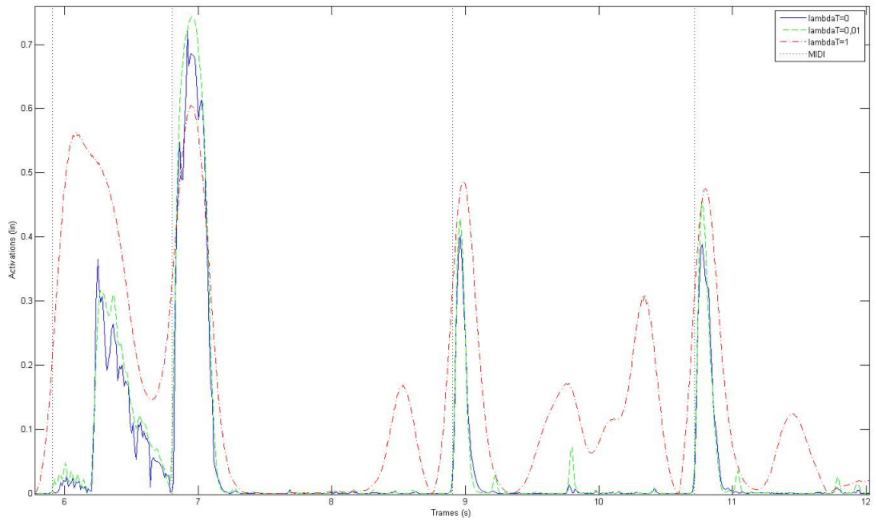


FIGURE 2.4 – Influence du poids de la contrainte de régularité sur la matrice d'activation (donnée pour une seule ligne de **H**). Le début des notes "vraies" est représenté par les lignes en pointillés verticales. On constate que la régularité effectue bien un lissage des activations, mais les activations parasites se trouvent amplifiées, puisqu'étendues.

avec n la dimension de $\vec{x}$. La parcimonie ainsi définie est égale à 1 si et seulement si une seule composante de $\vec{x}$ est non nulle, et à 0 si toutes les composantes de $\vec{x}$ sont égales, aussi utilisons-nous la fonction de coût suivante :

$$C_P(\mathbf{H}) = \frac{1}{\sqrt{R}-1} \sum_T \left[\frac{\sum_R |H_{rt}|}{\sqrt{\sum_R H_{rt}^2}} - 1 \right]. \tag{2.9}$$

C'est cette pénalité qui est implémentée dans l'ensemble des algorithmes testés, pondérée par un terme de poids λ_P. Un exemple est donné figure 2.5. On constate un soucis dans la polyphonie (à la fin de la séquence), où deux notes sont coupées puis reprises afin de satisfaire à la contrainte, ce qui ne traduit pas la réalité. Ainsi est-il nécessaire d'utiliser la contrainte de régularité pour palier à ce phénomène.

Nous proposons section suivante une paramétrisation en exponentielles décroissantes des lignes de **H**, qui sera évaluée en contrainte relaxée à l'image de la paramétrisation `IhR`

2.2.3 Paramétrisation de H sous forme d'exponentielles décroissantes

Dans le cadre de ce travail, nous avons proposé de réaliser l'approximation du spectrogramme par la factorisation suivante :

$$V_{ft} \sim \tilde{V}_{ft} = \sum_{r=1}^{R} W_{fr}^{\theta_r} \, H_{rt}^{\alpha_r} \tag{2.10}$$

avec

$$W_{fr}^{\theta_r} = \sum_{k=1}^{K_r} a_{r,k} \, g(f - f_{r,k}) \tag{2.11}$$

et

$$H_{rt}^{\alpha_r} = \sum_{i=1}^{I_t} A_{r,i} \, e^{-\delta_r (t - O_i)}. \tag{2.12}$$

Les paramétres du modèle de chaque ligne de **H** mis à jour à chaque itération sont :

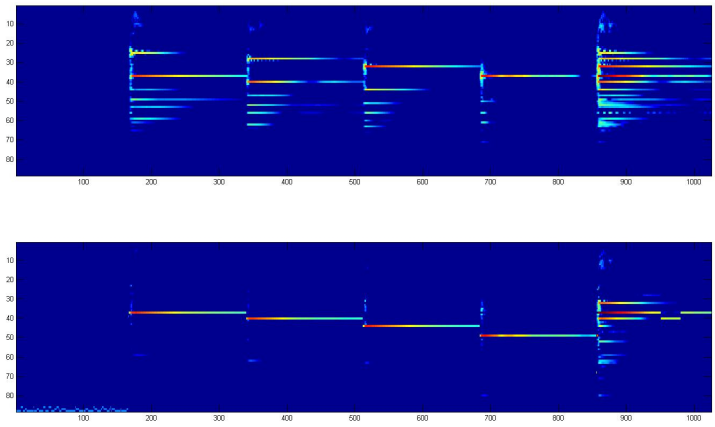


FIGURE 2.5 – Influence de la parcimonie sur la matrice d'activation, pour quatre notes jouées successivement puis simultanément. La figure du haut représente les activations de chaque atome sans parcimonie, la figure du bas avec un poids λ_P de 1. On constate que les activations parasites des notes isolées sont bien diminuées.

- $O_i,\ i \in [1, I]$, les onsets détectés dans toutes les lignes. Ils sont identiques pour tous les atomes.
- $A_{r,i}$, les amplitudes de chaque atome correspondant à l'onset i,
- δ_r, les coefficients de décroissance de chaque atome, identique pour tous les onsets.

Pour plusieurs raisons (notamment le fait que l'évolution de l'énergie rayonnée par un piano durant une note se fait suivant une double décroissance, *cf.* [Fletcher and Rossing, 1998]), cette contrainte ne peut être intégrée strictement. Elle sera donc intégrée de manière relaxée, ce qui présente l'inconvénient de nécessiter l'apprentissage du poids optimal pour cette contrainte (qui ne sera pas testée dans la suite, par manque de temps).

Une nouvelle fois, pour des raisons d'unification de la représentation (*cf.* [Raczyński et al., 2007], [Bertin, 2009]), on choisit de normaliser à chaque itération la matrice $\mathbf{H}$ par la valeur maximum pour obtenir $\max_{r,t}(h_{rt}) = 1$. De même, les amplitudes (a_k ou a_{rk}, selon la méthode) peuvent être normalisées, de telle sorte que pour chaque atome r_0, $\max_f(w_{fr_0}) = 1$. L'ensemble des variantes contraintes décrites ci-dessus seront évaluées chapître §4. Nous présentons au chapître suivant le système de transcription complet.

Chapitre 3

Système de transcription

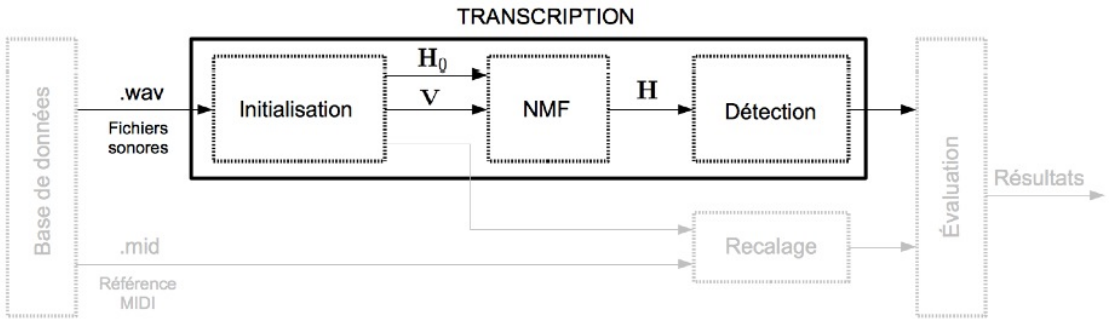


FIGURE 3.1 – Schématisation du système de transcription

Le système de transcription se décompose en trois parties (*cf.* figure 3.1). Il prend en entrée un fichier sonore au format *wav* mono (voie gauche dans le cas de signaux stéréos), et renvoie une estimation de l'activation de toutes les notes de la tessiture du piano, sous forme de liste (*cf.* § 4). Nous détaillons ici les procédures d'initialisation et de *post-traitement* qui encadrent les méthodes *NMF*.

3.1 Initialisation

La représentation temps-fréquence usuelle est la *Transformée de Fourier à court terme* (*TFCT*). Afin de se placer dans le cadre de la décomposition *NMF*, on peut choisir de ne travailler que sur le module sur spectre, ou sur le spectre de puissance du signal (tous deux assurément positifs). E. Vincent précise dans [Vincent et al., 2009] [§2.2] que l'utilisation du spectrogramme de puissance est plus adaptée à la modélisation d'une combinaison linéaire des différentes notes.

Cependant, dans la mesure où l'on souhaite pouvoir discriminer les fréquences fondamentales correspondantes aux notes de la gamme tempérée, une résolution minimum de un demi-ton sur toute la tessiture est souhaitée. Ceci peut être réalisé en jouant sur la durée de la fenêtre d'analyse, ce qui pénalise la détection des attaques (on remarquera qu'une précision de $1,5$ Hz est nécessaire pour disciminer les premiers partiels des deux notes les plus graves de la tessiture du piano, ce qui correspond à une fenêtre temporelle de $0.7s$ pour une fréquence d'échantillonnage de 44100 Hz). L'utilisation d'un banc de filtres dont

le facteur de qualité reste constant sur toutes les fréquences (*Constant Q Transform*) est une solution, de même que l'utilisation d'un banc de filtres qui approche la perception humaine : l'échelle *ERB*, pour *Equivalent Rectangular Bandwith* (*cf.*[Vincent et al., 2009]). Il est précisé que l'amélioration porte essentiellement sur le coût de calcul, et non sur les performances en terme de transcription.

Pour toute la suite, nous travaillons sur le module au carré de la *TFCT* calculée avec une fenêtre de *Hann* sur 4096 points, avec un incrément de 512 points sur les trames d'analyse. La fréquence d'échantillonnage de la base de données *MAPS* est fixée à 44100 Hz.

L'initialisation du dictionnaire **W** est choisie sur la base des paramètres d'accordage moyens estimés par François Rigaud au cours de sa thèse (donné dans [Rigaud et al., 2011], et rappelés figure 3.2). On trouvera dans [Bertin, 2009] §III.6 une liste d'articles qui se posent la question de l'initialisation de la matrice **W**, en laissant l'initialisation de **H** au hasard (ce qui rejoint l'idée de constituer un dictionnaire d'atomes). Les deux méthodes exposées dans [Bertin, 2009] sont basées sur un pré-clustering non supervisé de **V**. Les centroïdes des classes ainsi dégagées deviennent les colonnes de **W**.

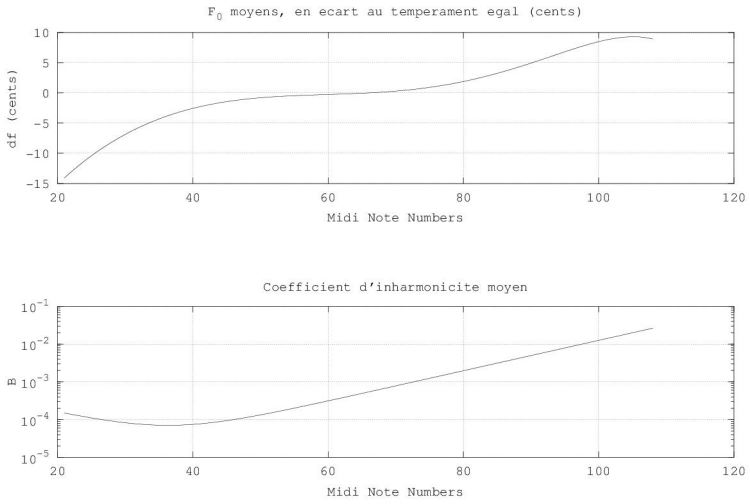


FIGURE 3.2 – Fréquences de référence et coefficients d'inharmonicité moyens utilisés pour l'initialisation du dictionnaire

On propose ici une méthode d'initialisation de la matrice d'activation **H** par une estimation des fréquences fondamentales contenues dans chaque trame. On se base sur les méthodes décrites dans [Galembo and Askenfelt, 1999] et [Klapuri, 2006], en remarquant qu'il en existe beaucoup d'autres (basées par exemple sur le principe de maximum de vraisemblance, *cf.* [Emiya et al., 2007]). On estime tout d'abord le niveau de bruit normalisé le long du spectrogramme au moyen d'un filtrage médian (*cf.* A), afin de limiter sa contribution à la détection des activations. La détection en elle même se réalise au moyen du *produit spectral inharmonique*, construit à partir de [Galembo and Askenfelt, 1999], puis d'une procédure de seuillage et normalisation.

3.1.1 Produit spectral inharmonique

On trouvera dans [Emiya, 2008] [§1.2.2] la définition du "*log-produit spectral* , qui peut être vu comme une somme des spectres compressés et exprimés en dB :

$$\begin{aligned} P(f) &= \sum_{h=1}^{H} 20 \log |X(hf)| \\ &= 20 \log \prod_{h=1}^{H} |X(hf)| \end{aligned} \tag{3.1}$$

La méthode proposée inclut la connaissance sur la distribution inharmonique des fréquences des partiels. Le filtre en peigne de dirac associé à l'atome r est

$$\gamma_r(f) = \sum_{k=1}^{K_r} \delta(f - f_{r,k}) \tag{3.2}$$

avec K_r (*ie.* $f_{r,K_r} < f_{max}$) donné par

$$K_r(f_{r,0}, B_r, F_e) = \left[\frac{\sqrt{1 + 4B_r \frac{f_{max}}{f_{r,0}}}}{2B_r} \right]^{1/2} \tag{3.3}$$

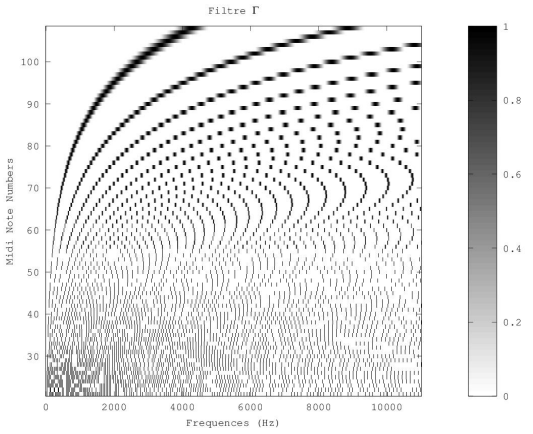


FIGURE 3.3 – Ensemble des filtres Γ_r pour $F_e = 44100$, $N_{fft} = 4096$, coupés à $F_e/4$

Le filtre de fréquence centrale $f_{r,k}$ est défini comme une fenêtre de *Hann* $\mathcal{W}$ de largeur L_w, qui est constante sur un atome et proportionnelle à la fréquence fondamentale. On choisit une largeur de 1/8 de ton. Le banc de filtre en amplitude de l'atome r est alors

$$\Gamma_r(f) = [\gamma_r \circledast \mathcal{W}](f) \tag{3.4}$$

où $\circledast$ dénote la convolution. On note $\mathcal{E}_r$ le facteur de normalisation de l'énergie contenue dans les bandes du filtre Γ_r

$$\mathcal{E}_r = \int |\Gamma_r(f)| \, df. \tag{3.5}$$

Ainsi, dans chaque trame, la puissance normalisée du signal contenue dans chaque filtre r est définie comme :

$$\begin{aligned} P(r) &= \frac{1}{\mathcal{E}_r} \sum_f \Gamma_r(f) \; 20 \log \frac{|\tilde{X}_s(f)|}{x_p(f)} \\ &= \frac{20}{\mathcal{E}_r} \log \prod_f \left(\frac{|\tilde{X}_s(f)|}{x_p(f)} \right)^{\Gamma_r(f)} . \end{aligned} \tag{3.6}$$

On note

– $\mathbf{X}$ le vecteur colonne des $F = N_{\text{fft}}/2$ premiers échantillons du résultat du filtrage médian normalisé par le niveau de bruit et exprimé en dB

$$\mathbf{X} = 20 \log \frac{|\tilde{X}_s(f)|}{x_p(f)} \tag{3.7}$$

– $\mathbf{P}$ le vecteur colonne des contributions des R atomes à $\mathbf{X}$,

– $\mathcal{E}$ le vecteur contenant l'ensemble des $\mathcal{E}_r$:

$$\mathcal{E} = \begin{bmatrix} \mathcal{E}_1 \\ \vdots \\ \mathcal{E}_R \end{bmatrix}. \tag{3.8}$$

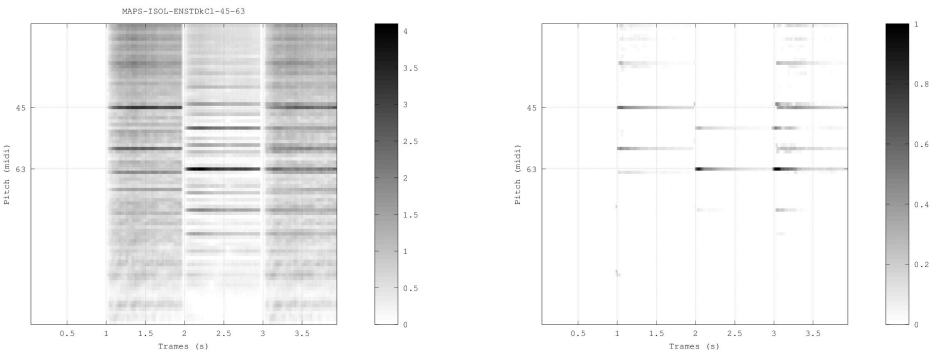
– $\mathbf{\Gamma}$ la matrice $R \times F$ définie par

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_1(1) & \cdots & \Gamma_1(f) & \cdots & \Gamma_1(F) \\ \vdots & & \vdots & & \vdots \\ \Gamma_r(1) & \cdots & \Gamma_r(f) & \cdots & \Gamma_r(F) \\ \vdots & & \vdots & & \vdots \\ \Gamma_R(1) & \cdots & \Gamma_R(f) & \cdots & \Gamma_R(F) \end{bmatrix}. \tag{3.9}$$

Alors

$$\mathbf{P} = \mathbf{\Gamma X} \oslash \mathcal{E} \tag{3.10}$$

où $\oslash$ dénote la division terme à terme (figure 3.4(a)).



(a) Log-produit spectral (equation 3.10) (b) Initialisation $\mathbf{H}_0$ (equation 3.14)

FIGURE 3.4 – Résultats du produit spectral *pré* et *post* seuillage, pour une séquence composée de deux notes jouées l'une après l'autre puis ensemble (La_2, midi 45 et $Mi_{\flat 4}$, midi 63, d'un piano DISKLAVIER tiré de la base *MAPS*) échantillonée à $44100Hz$, calculée avec une fenêtre de Hann de 2^{12} points et un incrément de 512 points sur les trames d'analyse, pour $\nu = 1$, $p = 0,999$, et une largeur de bande pour le calcul de la médiane de $300hz$.

3.1.2 Sélection et mise en forme

La fonction d'activation $\mathcal{P}$ est obtenue en sélectionnant les atomes pour lesquels $\mathbf{P}$ est supérieur à un certain seuil d'activation $\mathcal{S}$ paramétré par ν :

$$\mathcal{S} = 3\nu\sqrt{\text{Var}(\mathbf{P})} \tag{3.11}$$

alors

$$\mathcal{P} = \lfloor \mathbf{P} - \mathcal{S} \rfloor_0. \tag{3.12}$$

Après normalisation

$$\mathcal{P} = \frac{\mathcal{P}}{\max_r \mathcal{P}} \tag{3.13}$$

on pondère le résultat précédent par l'énergie normalisée du spectrogramme d'amplitude, afin de conserver l'information liée à l'énergie de chaque trame

$$\mathbf{H}_0 = \frac{\sqrt{\sum_f |X_0|^2}}{\max_t \sqrt{\sum_f |X_0|^2}} \mathcal{P} \tag{3.14}$$

Enfin, l'initialisation $\mathbf{H}_0$ de la matrice d'activation est obtenue en mettant à 10^{-1} (*ie.* $-20dB$) dans toutes les trames si l'atome est détecté activé au moins une fois au cours du morceau (figure 3.4(b)). La matrice $\mathbf{H}_0$ constitue une matrice d'activation en soi. Les performances de transcription obtenues pour ce système d'estimation (noté `PS`) de base seront comparées à celles des méthodes *NMF* plus complexes décrites §2. L'algorithme écrit est donné ci après :

Algorithme d'initialisation

INPUT: Fichier .wav
OUTPUT: Spectrogramme normalisé $\mathbf{V}$ et matrice d'activation $\mathbf{H}_0$

Générer la matrice $\mathbf{\Gamma}$ et le vecteur $\mathcal{E}$ (Eq.3.9)
Initialiser la *TFct* $\mathbf{X}$ et la matrice $\mathbf{H}_0$
Pour toutes les trames
    Calcul de la *TFD* de la trame $\longrightarrow$ $\mathbf{X}$
    Estimation du niveau de bruit (Eq. A.6)
    Calcul de $\mathbf{P}$ (Eq. 3.10)
    Traitement de $\mathbf{P}$ (Eq. 3.13)
fin Pour
Calcul du spectrogramme $\mathbf{V} \longleftarrow |\mathbf{X}|^2 \; \forall \; f \; \in \; [0 : F_e/2 \,]$
Normalisation $\mathbf{V} \longleftarrow \mathbf{V} / \max_{ft}(\mathbf{V})$
Pondération de $\mathbf{H}_0$ (Eq. 3.14)

3.2 Post-traitement

L'initialisation obtenue précédemment est utilisée pour chaque variante contrainte *NMF*, qui sont décrites §2. Cependant, la matrice $\mathbf{H}$ des activations dans le temps de chaque atome ne constitue pas un résultat de transcription en soi. En effet, l'objectif est d'extraire les caractéristiques des *notes* en présence. Nous nous limiterons dans cette étude à la détermination de l'**amplitude** (nuance, ou *vélocité*), des temps de **début** (*onset*) et de **fin** (*offset*), pour chaque note repérée par sa **hauteur** (*pitch*). On remarquera que, dans le cas d'un dictionnaire laissé libre (*cf.* [Vincent et al., 2007]), il est nécessaire d'évaluer la

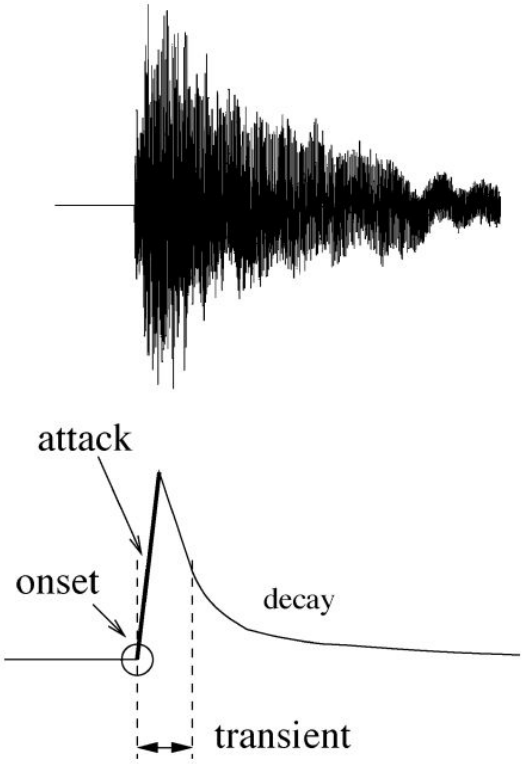


FIGURE 3.5 – Onset, attaque, transitoire et déclin dans le cas idéal d'une note isolée (tiré de [Bello et al., 2005]).

hauteur perçue pour chaque atome.

Cependant, la tâche n'est pas simple. La définition même d'un onset varie d'un auteur à l'autre, et celle d'un offset est généralement passée sous silence (pour la hauteur, la correspondance avec la fréquence est immédiate). On trouvera cependant dans [Raczyński et al., 2007] une méthode de détection de note pour la transcription par *NMF* qui donne à la fois les *onsets* et les *offsets*. J.P. Bello propose dans [Bello et al., 2005] de définir un onset comme le *début de l'attaque* d'une note (*cf.* figure 3.5). Une méthode répandue pour la détection des onsets consiste alors à détecter une brusque croissance dans l'enveloppe énergétique d'un signal.

3.2.1 Lissage et différenciation des activations

La piste d'activation présente de nombreuses variations (battements), *a priori* indépendantes des amplitudes spectrales d'origine (*cf.* 3.6). Aussi est-il décidé de *lisser* chaque ligne de $\mathbf{H}$ par convolution avec un *filtre d'oubli*, modélisé par un filtre auto-régressif à l'ordre 1, de coefficient $a = -0.95$.

Pour le piano, instrument dit percussif, la détection d'onsets se fait en repérant les brusques croissances dans les lignes de $\mathbf{H}$. Nous nous intéressons donc à la dérivée temporelle de chaque ligne. Le calcul de la dérivée est réalisé au moyen d'un filtre différenciateur, généré par la *méthode de Rémez* (*cf.* [McClellan and Parks, 1973]). La différence avec un filtre

différenciateur standard (de noyau de convolution $[1 - 1]$) est que le filtre utilisé est moins sensible aux hautes fréquences, soit aux variations rapides de la piste d'activation (ce qui en fait un complément à l'étape de lissage précédente).

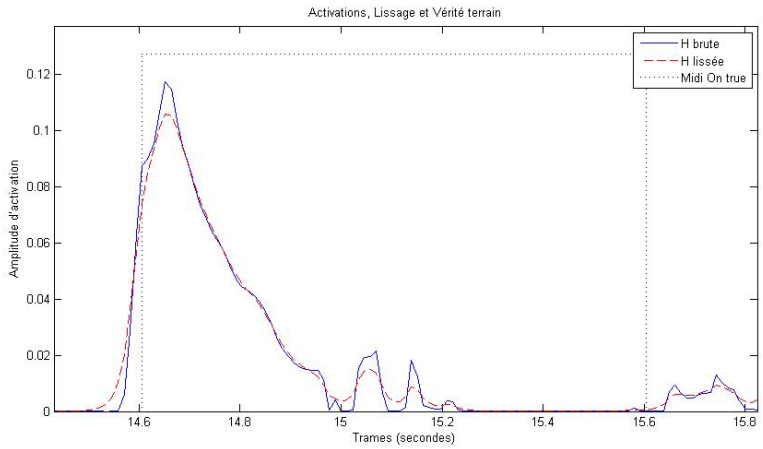


FIGURE 3.6 – Sortie brute de NMF et son lissage, comparés à la vérité terrain.

3.2.2 Détection des onsets/offsets

La détection d'onsets (respectivement d'offsets) se fait en posant un seuil dans la partie positive (respectivement négative) de la différenciation de chaque ligne. Le passage d'une valeur inférieure à une valeur supérieure au seuil de détection d'onsets marque le début de l'attaque ; de même que le passage d'une valeur inférieure à une valeur supérieure au seuil de détection d'offsets marque la fin de la décroissance d'une note. L'onset est alloué au *bin* temporel correspondant au maximum de la différenciation entre le début de l'attaque et la fin de la décroissance. L'offset est aloué au *bin* correspondant à la fin de la décroissance (*cf.* 3.7).

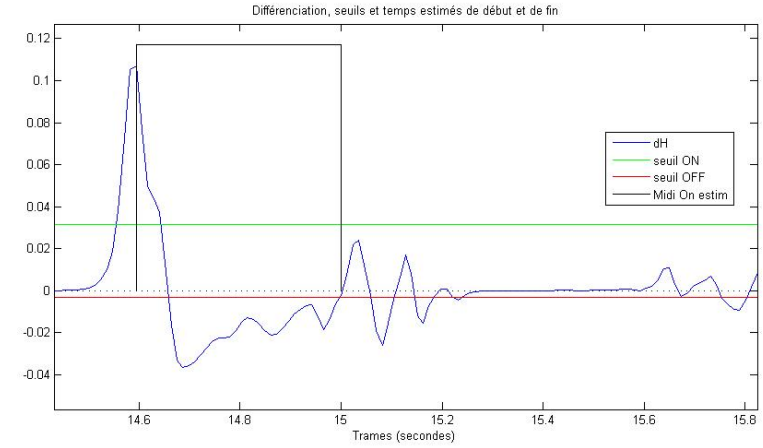


FIGURE 3.7 – Différentiation de **H**, seuils de detection des onsets et des offsets, et estimation de la note.

3.2.3 Regroupement

Pour prendre en compte les limitations physiologiques en terme de rapidité de répétition d'une note, il est supposé qu'aucune pièce musicale ne nécessite de répéter plus de 10 fois par seconde la même note. Pour chaque note qui en suit une autre (*ie.* à l'exception de la première), on calcule la durée entre les deux onsets. S'ils sont distants de moins de $100\,ms$, les notes sont regroupées ; c'est à dire remplacées par une seule note dont l'onset est celui de la première et l'offset celui de la dernière. L'opération est répétée tant qu'un regroupement est possible, et pour chaque ligne de $\mathbf{H}$.

La liste de *notes* ainsi obtenue est comparée à la vérité terrain, pour obtenir les performances de transcription de chaque variante *NMF* testée ; ceci est décrit au chapitre suivant. Nous détaillons maintenant le choix du réglage des paramètres du système.

3.3 Réglage des paramètres

L'ensemble de la procédure de transcription nécessite l'apprentissage des valeurs optimales pour un grand nombre de paramètres. Ceci est réalisé sur les 10 premières secondes d'un corpus de 10 morceaux choisis au hasard dans la base *MAPS* (voir le détail de la base de développement en annexe D.2). Le critère retenu est la $\mathcal{F}$-mesure, décrite dans la procédure d'évaluation §4.2. Ainsi, les performances de chaque système sont évaluées pour chaque réglage, afin de choisir les paramètres qui maximisent le critère. Nous donnons ci-après les paramètres retenus pour chaque partie du système proposé.

3.3.1 Paramètres pour l'initialiation

Les paramètres de la TFCT sont choisis sur la base des remarques faites §3.1. L'ensemble des paramètres est appris sur la base de développement, et est donné table 3.1.

Paramètre	Valeur retenue
Ordre de la TFD	4096 points
Pas d'incrémentation	512 points
Largeur de la fenêtre pour le filtrage médian	300 Hz
Ordre du filtrage médian	0.9999 percentiles
$F_{\max}$ pour le calcul du produit spectral	10 kHz
seuil de détection ν	1

TABLE 3.1 – Paramètres choisis pour la phase d'initialisation.

3.3.2 Paramètres pour les méthodes *NMF*

Comme expliqué §2.2, l'estimation des poids optimaux pour les contraintes de régularité (§2.2.1) et de parcimonie (§2.2.2) est réalisée conjointement. On démontre ainsi qu'elles ont une influence l'une sur l'autre, puisque les poids sont choisis (proches mais) différents de zéro (*cf.* figure 3.8).

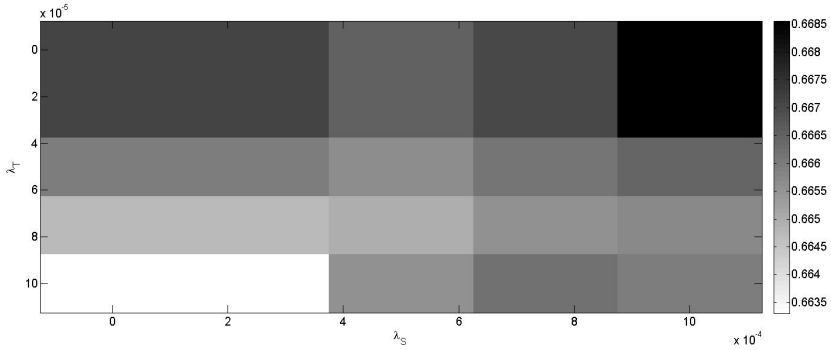


FIGURE 3.8 – Réglage des poids respectifs des contraintes de régularité λ_P (en abscisse) et de parcimonie λ_P (en ordonnée) en maximisant la $\mathcal{F}$-mesure sur la base de développement, pour la méthode `IhR`. On choisit $\lambda_T = 2.10^{-5}$ et $\lambda_P = 10^{-3}$, pour un score de 66,85%.

Méthode	`Ha`	`Ih`	`IhR`
Nombre d'itérations	50		
Tessiture couverte (Midi note numbers)	21 à 108		
$F_{\max}$ des partiels de chaque atome	10 kHz		
Nombe maximum de partiels de chaque atome	15		
Ordre de la β-divergence de la fonction de coût principale	1		
Ordre de la β-divergence de la fonction de coût auxiliaire	•		2
Poids de la fonction de coût auxiliaire	•		10^{-2}
Poids de la contrainte de régularité	2.10^{-5}		
Poids de la contrainte de régularité	10^{-3}		

TABLE 3.2 – Paramètres choisis pour les méthodes *NMF*

3.3.3 Paramètres pour la détection

Un exemple est donné figure 3.9, où l'on présente l'évolution de la $\mathcal{F}$-mesure avec le seuil de détection des onsets (§3.2.2) pour les quatre méthodes (`PS`, `Ha`, `Ih` et `IhR`). On choisit ainsi -15 dB pour le produit spectral (l'initialisation $\mathbf{H}_0$) et -30 dB pour les trois méthodes NMF[1]. Tous les parmètres sont donnés en table 3.3.

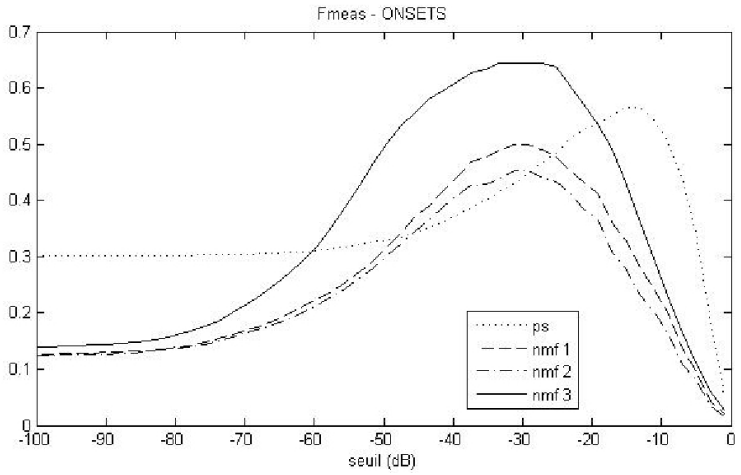


FIGURE 3.9 – Réglage du seuil de détection des onsets en maximisant la $\mathcal{F}$-mesure sur la base de développement. Les acronymes `nmf1`, `nmf2` et `nmf3` correspondent respectivement aux méthodes `Ha`, `Ih` et `IhR`.

Méthode	`PS`	`Ha`	`Ih`	`IhR`
Coefficient a pour le lissage	-0.95			
Seuil de détection des onsets	-15 dB	-30 dB		
Seuil de détection des offsets	-80 dB			

TABLE 3.3 – Paramètres choisis pour la détection

Nous présentons au chapitre suivant la procédure d'évaluation et les résultats obtenus.

1. On remarquera que la méthode `IhR` (`nmf3` sur la figure 3.9) atteint des performances de l'ordre de 65% sur la base de développement.

Chapitre 4

Évaluation

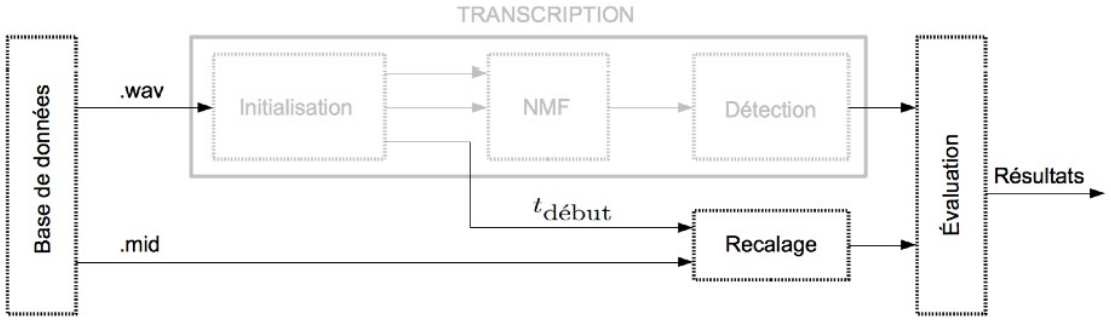


FIGURE 4.1 – Schématisation du système de test dans son ensemble.

4.1 Base de données

Afin d'évaluer précisément les résultats de transcription, il est nécessaire de disposer de sons de qualité et d'une référence précise. Nous utiliserons donc la base de données *MAPS* (pour "*MIDI Aligned Piano Sounds*") développée par Valentin Emiya dans le cadre de sa thèse [Emiya, 2008]. Cette base se compose de plusieurs morceaux du répertoire de musique pour piano, transcrits au format MIDI sous forme de fichiers *SMF* ("*Standard MIDI File*"), disponibles sur le site internet de B.Krueger [1] sous license *Creative Commons*. On notera que la place, la durée et l'intensité de chaque note ont été ajustées par l'auteur (l'interprète).

Pour chaque instrument et condition d'enregistrement (*cf.* table D.1), 30 morceaux ont été choisis au hasard. Les fichiers son (au format *".wav"*) ont ensuite été générés pour chaque fichier *".smf"* suivant deux procédés :

- Pour le piano Disklavier de Yamaha, les fichiers *".smf"* sont joués, et le son rayonné par l'instrument est enregistré à deux distances (prise de proximité *Close*, et prise d'ambiance *Ambient*).
- Pour les synthétiseurs logiciels, la diversité des conditions d'enregistrement se traduit par l'utilisation de réverbérations synthétiques différentes

1. B. Krueger : Classical Piano MIDI files. http ://www.piano-midi.de, 2008.

L'évaluation sur l'ensemble de la base n'étant pas réalisable pour des questions de temps de calcul, nous sélectionnons 15 morceaux, en tentant de couvrir l'ensemble des performances atteintes par la méthode développée par V. Emiya, donné sur sa page web [2] et nous nous limiterons aux 30 premières secondes de chaque morceau choisi (*cf.* D.3). Disposant de fichiers *.wav* et de leurs correspondances en *.mid*, on choisit la vérité terrain suivante :

– Estimation du temps de début du morceau à partir de l'énergie normalisée du signal S (*.wav*) :

$$t_{\text{début}} = \min_t \mid 20 \log \left(\frac{S(t)^2}{\max_t S(t)^2} \right) > -50 \text{ dB}. \tag{4.1}$$

– Extraction des informations du fichier *.mid*, rangées dans une matrice dont chaque ligne n correspond à une note, et est constituée des données midi usuelles : piste, canal, numéro de note, vélocité, onset, offset, message1, message2. ceci est réalisé au moyen de la toolbox "Matlab Midi" developpée par Ken Schutte, disponible sur sa page web [3].

– Décalage en bloc de chaque note d'un temps correspondant à $t_{\text{début}}$, et création d'une liste de notes dont chaque ligne est au format [pitch t_{on} t_{off}].

Le décalage est nécessaire pour s'assurer que les fichiers *midi* et *audio* sont bien synchronisés. On vérifie ensuite que les fichiers sont interprétés à la même vitesse.

4.2 Critères d'évaluation

4.2.1 $\mathcal{F}$-mesure

Dans la mesure où l'on ne s'intéresse pas à l'estimation de l'amplitude de chaque note, les évaluations correspondent à une classification binaire (item détecté ou non-détecté) [4]. Ces scores sont basés sur le comptage du nombre de vrais et de faux, positifs et negatifs (*cf.* table 4.1).

item	Présent	Non-présent
Détecté	VP	FP
Non-détecté	FN	VN

TABLE 4.1 – Critères d'évaluation binaire

On trouvera dans [Rijsbergen, 1979] l'expression de fonctions d'évaluation des performances de tels descripteurs, couramment utilisées en évaluation des performances de transcription : la *précision* $\mathcal{P}$, le *rappel* $\mathcal{R}$ et une combinaison des deux, la *F-mesure* $\mathcal{F}$.

2. V. Emiya, Automatic transcription of piano music, http ://www.lif.univ-rs.fr/ vemiya/EUSIPCO08/results.html, 2008

3. K. Schutte, MATLAB and MIDI, http ://www.kenschutte.com/midi, 2012

4. On remarquera que Adrien Daniel *et al.* présentent dans [Daniel et al., 2008] une évaluation du poids perceptif des erreurs typiques commises pas les systèmes de transcription automatique de musique polyphonique. En terme de hauteur, les erreurs pour lesquelles on dispose d'une pondération perceptive sont les erreurs d'octave et de quinte (tous les autres intervalles sont affublés du même poids perceptif). Ensuite viennent les importances des erreurs de durée et d'onset.

– Rappel : représente la capacité du système à trouver les bonnes notes (de 0 pour aucune note trouvée à 1 si toutes les notes sont trouvées)

$$\mathcal{R} = \frac{VP}{VP + FN} \tag{4.2}$$

– Précision : représente la capacité du système à ne trouver que les bonnes notes (vaut 1 si uniquement des bonnes notes ont été trouvées et diminue si moins de bonnes ou plus de fausses notes sont trouvées)

$$\mathcal{P} = \frac{VP}{VP + FP} \tag{4.3}$$

– $\mathcal{F}$-mesure : un compromis entre la précision et le rappel

$$\mathcal{F} = 2\frac{\mathcal{P}\mathcal{R}}{\mathcal{P} + \mathcal{R}} \tag{4.4}$$

L'attention est portée sur les résultats en termes de $\mathcal{F}$-mesure, comme il est d'usage dans les campagnes d'évaluations.

4.2.2 *Mean Overlap Ratio*

Ryynänen et Klapuri proposent dans [Ryynänen and Klapuri, 2005] un critère d'évaluation basé à la fois sur l'estimation des onsets et sur le rapport entre la durée estimée (transcrite) et réelle d'une note : le *rapport de recouvrement* ("*overlap ratio*" dans le texte), défini comme le rapport entre la durée de l'intersection et celle de l'union. Pour la note n,

$$o_n = \frac{\min_t\{\text{offset}_{\text{ref}}, \text{offset}_{\text{est}}\} - \max_t\{\text{onset}_{\text{ref}}, \text{onset}_{\text{est}}\}}{\max_t\{\text{offset}_{\text{ref}}, \text{offset}_{\text{est}}\} - \min_t\{\text{onset}_{\text{ref}}, \text{onset}_{\text{est}}\}} \tag{4.5}$$

avec $\text{onset}_{\text{ref/est}}$ (respectivement $\text{offset}_{\text{ref/est}}$) les temps de début (respectivement d'extinction) de la note vraie et transcrite. Le critère global d'évaluation pour un morceau est obtenu en moyennant sur l'ensemble des notes les rapports de recouvrements, définissant ainsi le *Mean Overlap Ratio*, abrégé en "*MOR*" :

$$MOR = \frac{1}{N}\sum_{n=1}^{N} o_n \tag{4.6}$$

4.3 Méthode

L'algorithme qui réalise l'évaluation prend en entrée deux listes de notes correspondantes à la référence et l'estimation au format [*pitch time_on time_off*] , ainsi qu'une durée de tolérance d_{tol} sur l'estimation des onsets, et renvoie les résultats de *precision*, *recall*, *f-mesure* et *MOR*, ainsi qu'une liste des notes oubliées et une liste des notes ajoutées lors de la transcription.

Le programme se résume comme suit : Pour chaque note n de la liste de référence, recherche d'une note de même *pitch* dont l'onset se situe dans $[time\_on_{\text{ref}} - d_{\text{tol}}, time\_on_{\text{ref}} + d_{\text{tol}}]$

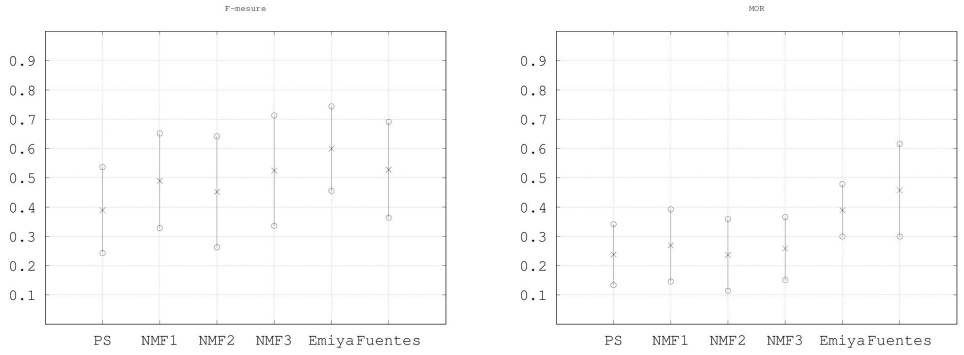
– si un onset est trouvé → calcul du *MOR*, suppression des lignes correspondantes des listes de référence et estimée, et $\#VP = \#VP + 1$;
– si aucun onset est trouvé → $\#FN = \#FN + 1$ et le taux de recouvrement pour cette note est posé égal à zéro ;

Le nombre de faux positifs est alors égal au nombre de notes restantes dans la liste estimée, qui devient la liste de note ajoutées ; la liste de notes oubliées est la liste de notes restantes dans la liste de référence.

4.4 Résultats

Les performances de transcription des algorithmes proposés seront comparées à celles de deux méthodes tirées de l'état de l'art. Tout d'abord, un système de transcription développé par Valentin Émiya dans le cadre de sa thèse, basé sur un système de chaine de Markov (*cf.* [Emiya et al., 2008], sous l'acronyme `Emiya`), ainsi qu'un système développé par Benoit Fuentes actuellement en poste de doctorant dans l'équipe AAO de Télécom Paristech, basé sur la *PLCA* (pour *"Probabilistic Latent Component Analysis", cf.* [Fuentes et al., 2011], sous l'acronyme `Fuentes`).

Afin d'être totalement objectif, nous donnerons le temps mis par chaque système pour réaliser la transcription, ramené sur la durée effective du corpus de morceaux de la base de test.

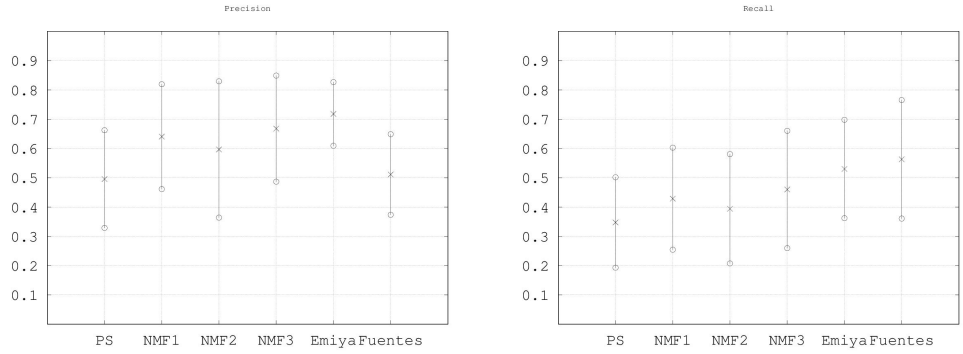


(a) $\mathcal{F}$-mesure (b) *Mean Overlap Ratio*

Figure 4.2 – Représentation graphique des résultats obtenus en termes de $\mathcal{F}$-mesure et de *Mean Overlap Ratio*, pour tous les systèmes de transcription, sur la base de test décrite table D.3

Tout d'abord, en termes de $\mathcal{F}$-mesure, on peut constater figure 4.2(a) que le système `PS` (constitué de la seule initialisation de la matrice $\mathbf{H}_0$ et du *post-traitement*) apporte plus de la moitié des performances de transcription. Ensuite, l'apport du modèle d'inharmonicité proposé par François Rigaud amène, dans sa version relaxée, une amélioration des performances comparées à celles de la méthode harmonique (initialement proposée par Romain Hennequin) de l'ordre de 5%. Le système qui présente les meilleurs résultats est celui proposé par Valentin Émiya, suivi de près par les système `IhR` (pour *"NMF inharmonique relaxée"*) et proposé par Benoit Fuentes, qui sont presque aussi performants.

Cependant, on constatera que, sur les notes détectées, la méthode proposée par Benoit Fuentes est mieux à-même de transcrire la durée de tenue des notes, ce que l'on retrouve figure 4.2(b). L'ensemble des systèmes basés sur les méthodes *NMF* ont des scores de *MOR* en-dessous de l'état de l'art, ce qui est dû à la procédure de *post-traitement*.



(a) Précision (b) Rappel

FIGURE 4.3 – Représentation graphique des résultats obtenus en termes de précision et de rappel, pour tous les systèmes de transcription, sur la base de test décrite table D.3

Plus précisément, on peut voir figure 4.3 que le système `IhR` est plus performant en termes de rappel et de précision que la méthode harmonique `Ha`. La méthode de Benoit Fuentes retrouve plus de bonnes notes (rappel plus élevé), mais renvoie aussi plus de fausses notes (précision plus faible) que le système `IhR`.

Le détail des résultats donné tableau 4.2 met en évidence la compétitivité des systèmes de transcription à base de méthode *NMF*. Car, si la méthode présentée par Valentin Émiya apporte les meilleurs résultats, elle nécessite $99,1$ secondes de calcul pour 1 seconde de morceau. Ainsi, le ratio performance/temps de calcul est le plus élevé pour la méthode de Benoit Fuentes, ensuite pour la méthode `IhR`, puis pour la méthode de V. Émiya.

Méthode	Précision		Rappel		$\mathcal{F}$-mesure		*MOR*		Temps
	Moy.	Éc.	Moy.	Éc.	Moy.	Éc.	Moy.	Éc.	
`PS`	49.5	16.6	34.7	15.4	38.9	14.6	23.7	10.5	10^{-6}
`Ha`	64.0	17.9	42.8	17.4	48.9	16.1	26.8	12.3	26.7
`Ih`	59.6	23.2	39.4	18.6	45.2	18.9	23.6	12.2	156.8
`IhR`	66.7	18.1	45.9	20.0	52.4	18.8	25.8	10.7	36.8
`V. Émiya`	71.7	10.8	53.0	16.8	59.9	14.4	38.9	9.0	99.1
`B. Fuentes`	51.1	13.7	56.3	20.2	52.7	16.3	45.7	15.8	0.44

TABLE 4.2 – Détail des résultats obtenus pour la système de base, les trois systèmes de transcription par *NMF* et les deux méthodes de référence, sur la base de test décrite table D.3

Conclusions

L'objectif du présent document est l'évaluation des performances de transcription d'algorithmes basés sur la factorisation du spectrogramme en matrices non-négatives, et de l'apport des modèles d'inharmonicité pour la tâche de transcription aveugle de musique polyphonique pour piano solo.

Les différentes variantes contraintes de la *NMF* ont tout d'abord été implémentées dans un cadre unifié (même représentation des données, même *pre/post-traitement*). Après une phase d'apprentissage des paramètres optimaux pour chaque méthode, l'ensemble des algorithmes a été testé sur un corpus de 15 morceaux et comparé à deux systèmes de transcription représentatifs des performances de l'état de l'art.

Les procédures d'initialisation et de *post-traitement* développées renvoient à elles seules des résultats de transcription dont les performances sont de l'ordre de 60% de l'état de l'art.

Il est ressorti que le modèle de distribution des fréquences des partiels de chaque notes suivant la loi d'inharmonicité considérée (§1.2) amène une amélioration de l'ordre de 10% des performances, par rapport à la version harmonique. On notera que la mise à jour des paramètres est plus efficace dans le cas où la contrainte est posée sous forme de pénalité (contrainte relaxée), et non en forme générale de la matrice $\mathbf{W}$ (contrainte stricte).

Les performances de la méthode retenue (`IhR`) sont comparables à l'état de l'art, en terme d'exactitude de la transcription et de temps de calcul. Ceci est prometteur, dans la mesure où l'on peut identifier, *via* les différents critères d'évaluation, les points du sytème à améliorer.

Ainsi, il peut être intéressant de travailler sur la procédure de détection des onsets, pour améliorer le rappel de la méthode. Puisque le contenu inharmonique d'une note de piano ne se révèle qu'après le transitoire d'attaque, l'ajout d'un modèle de composantes bruitées dans le système, qui servirait à la détection des onsets, est une solution (le modèle ARMA proposé dans [Hennequin et al., 2011] a été envisagé).

Les tests perceptifs de rendu de la décomposition *NMF* ont amené la certitude que toutes les notes en présence sont détectées, et que plusieurs atomes sont activés pour expliquer une seule note (typiquement les octaves et quintes). Ainsi, l'ajout d'un modèle probabiliste de note tel que présenté dans [Ryynänen and Klapuri, 2005] est une solution. Il a été envisagé d'appliquer cette méthode à une série de douzes matrices dérivées de $\mathbf{H}$, correspondantes aux douzes notes de la gamme tempérée (à l'image de *chromas*). On peut ainsi espérer améliorer grandement la précision de la méthode.

De plus, les résultats en termes de *taux moyen de recouvrement* (*MOR*) sont relativement faibles pour tous les systèmes proposés. Ceci provient de la procédure de détection des offsets. Un simple seuil dans la partie négative de la différenciation de chaque ligne de **H** est insuffisant, puisque très sensible aux battements évoqués §3. Une nouvelle fois, un modèle de note s'avère nécessaire pour relever ce score.

La transcription aveugle de musique polyphonique pour instrument solo par factorisation automatique du spectrogramme en matrices non-négatives incluant des contraintes liées à la connaissance de la physique de l'instrument est donc une voie prometteuse à explorer pour de futurs travaux.

Bibliographie

[Bello et al., 2005] Bello, J., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., and Sandler, M. (2005). A tutorial on onset detection in music signals. *Speech and Audio Processing, IEEE Transactions on*, 13(5) :1035 – 1047.

[Bertin, 2009] Bertin, N. (2009). *Factorisations en matrices non-négatives. Approches contraintes et probabilistes, application à la transcription automatique de musique polyphonique*. PhD thesis, École Nationale Supérieure des Télécommunications.

[Cichocki et al., 2009] Cichocki, A., Phan, A. H., and Zdunek, R. (2009). *Nonnegative Matrix and Tensor Factorizations : Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley, Chichester.

[Daniel et al., 2008] Daniel, A., Emiya, V., and David, B. (2008). Perceptually-based evaluation of the errors usually made when automatically transcribing music. In *Proc. Int. Conf. Music Information Retrieval (ISMIR)*, Philadelphia, PA, États-Unis.

[Durrieu, 2010] Durrieu, J.-L. (2010). *Transcription et séparation automatique de la mélodie principale dans les signaux de musique polyphoniques*. These, Télécom ParisTech.

[Emiya, 2008] Emiya, V. (2008). *Transcription automatique de la musique de piano*. PhD thesis, Télécom ParisTech.

[Emiya et al., 2007] Emiya, V., Badeau, R., and David, B. (2007). Multipitch estimation of quasi-harmonic sounds in colored noise. In *10th Int. Conf. on Digital Audio Effects (DAFx-07)*, Bordeaux, France. ANR-06-JCJC-0027 ; Music Discover ACI-Masse de données.

[Emiya et al., 2008] Emiya, V., Badeau, R., and David, B. (2008). Automatic transcription of piano music based on HMM tracking of jointly-estimated pitches. In *Proc. Eur. Conf. Sig. Proces. (EUSIPCO)*, Lausanne, Suisse. FP6-027026-K-SPACE.

[Fletcher and Rossing, 1998] Fletcher, N. and Rossing, T. (1998). *The Physics of Musical Instruments*. Springer.

[Fuentes et al., 2011] Fuentes, B., Badeau, R., and Richard, G. (2011). Adaptive harmonic time-frequency decomposition of audio using shift-invariant PLCA. In *Proc. of ICASSP*, pages 401–404, Prague, Czech Republic.

[Galembo and Askenfelt, 1999] Galembo, A. and Askenfelt, A. (1999). Signal representation and estimation of spectral parameters by inharmonic comb filters with application to the piano. *IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING*, 7(2) :197–203.

[Hennequin et al., 2010] Hennequin, R., Badeau, R., and David, B. (2010). Time-dependent parametric and harmonic templates in non-negative matrix factorization. In *International Conference On Digital Audio Effects*, pages 246–253, Graz, Austria.

[Hennequin et al., 2011] Hennequin, R., Badeau, R., and David, B. (2011). Nmf with time&#x2013 ;frequency activations to model nonstationary audio events. *Trans. Audio, Speech and Lang. Proc.*, 19(4) :744–753.

[Hoyer, 2004] Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.*, 5 :1457–1469.

[Klapuri, 2006] Klapuri, A. (2006). Multiple fundamental frequency estimation by summing harmonic amplitudes. In *in ISMIR*, pages 216–221.

[Lee and Seung, 1999] Lee, D. D. and Seung, S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401 :788–791.

[McClellan and Parks, 1973] McClellan, J. and Parks, T. (1973). A united approach to the design of optimum fir linear-phase digital filters. *Circuit Theory, IEEE Transactions on*, 20(6) :697 – 701.

[Moorer, 1977] Moorer, J. A. (1977). On the transcription of musical sound by computer. *Computer Music Journal*, 1(4) :pp. 32–38.

[Olshausen and Field, 1997] Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an overcomplete basis set : A strategy employed by v1 ? *Vision Research*, 37(23) :3311 – 3325.

[Paatero and Tapper, 1994] Paatero, P. and Tapper, U. (1994). Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5(2) :111–126.

[Raczyński et al., 2007] Raczyński, S. A., Ono, N., and Sagayama, S. (2007). Multipitch analysis with harmonic nonnegative matrix approximation. In *in ISMIR 2007, 8th International Conference on Music Information Retrieval*, pages 381–386.

[Rigaud et al., 2011] Rigaud, F., David, B., and Daudet, L. (2011). A parametric model of piano tuning. In *Proc. of the 14th Int. Conf. on Digital Audio Effects (DAFx-11)*, pages 393–399.

[Rijsbergen, 1979] Rijsbergen, C. J. V. (1979). *Information Retrieval*. Butterworth-Heinemann, Newton, MA, USA, 2nd edition.

[Ryynänen and Klapuri, 2005] Ryynänen, M. P. and Klapuri, A. (2005). Polyphonic music transcription using note event modeling.

[Smaragdis, 2003] Smaragdis, P. (2003). Non-negative matrix factorization for polyphonic music transcription. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2003*, pages 177–180.

[Vincent et al., 2007] Vincent, E., Bertin, N., and Badeau, R. (2007). Two nonnegative matrix factorization methods for polyphonic pitch transcription. In *2007 Music Information Retrieval Evaluation eXchange (MIREX)*, Vienna, Autriche.

[Vincent et al., 2008] Vincent, E., Bertin, N., and Badeau, R. (2008). Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription. In *2008 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 109–112, Las Vegas, États-Unis.

[Vincent et al., 2009] Vincent, E., Bertin, N., and Badeau, R. (2009). Adaptive harmonic spectral decomposition for multiple pitch estimation. Rapport de recherche PI 1919. This technical report is deprecated. Please refer to the following article instead : http ://hal.inria.fr/inria-00544094/.

[Virtanen, 2004] Virtanen, T. (2004). Separation of sound sources by convolutive sparse coding. In *in Proc. ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing, 2004. [Online] Available : http ://journal.speech.cs.cmu.edu/SAPA2004*.

[Virtanen, 2007] Virtanen, T. (2007). Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria. *IEEE Transactions on Audio, Speech and Language Processing*, 15(3) :1066–1074.

[Yeh and Röbel, 2006] Yeh, C. and Röbel, A. (2006). Adaptive noise level estimation. In *in Workshop on Computer Music and Audio Technology(WOCMAT'06*.

ANNEXES

Annexe A

Filtrage Médian

On suppose que le signal de bruit x_b est additif au signal utile x_s , et est le résultat du filtrage d'un processus aléatoire gaussien de moyenne nulle (bruit coloré). Par linéarité de la Transformée de Fourier discrète, on peut écrire

$$X_0(f) = X_s(f) + X_b(f)(1) \tag{A.1}$$

avec X_s le spectre correspondant aux partiels en présence, et X_b le spectre du bruit. Afin de pouvoir définir un seuil unique de détection indépendamment du niveau du signal, on considère le spectrogramme normalisé par l'énergie qu'il contient. Ainsi, on travaille sur $|X|$ défini par :

$$|X| = \frac{|X_0|}{\sqrt{\sum_f |X_0|^2}} \tag{A.2}$$

avec $|X_0|$ le spectre observé. Le niveau de bruit coloré est défini comme l'espérance sur l'amplitude des observations des pics de bruit, avec comme définition "un pic qui ne peut s'interpréter comme issu d'un processus sinusoïdal, stationnaire ou modulé" (*cf.* [Yeh and Röbel, 2006]). Le niveau de bruit coloré peut alors être interprété comme une approximation suivant f du niveau de bruit moyen, et représenté par une courbe à variations lentes.

Les observations $|X|$ sont les valeurs du module de la TFD du signal dans une bande fréquentielle centrée sur f, suffisamment étroite pour que la contribution des pics correspondants aux partiels soit négligeable. On suppose que les observations suivent une distribution de Rayleigh $p_{\mathcal{R}}$ (de support semi-infini) de mode σ (valeur la plus fréquemment observée) définie par

$$p_{\mathcal{R}}(x, \sigma) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}; \quad x \in [0, \infty], \ \sigma \in \mathbb{R}_+ \tag{A.3}$$

de moyenne, variance et médiane (respectivement)

$$\begin{aligned} \mathbb{E}[X] &= \sigma\sqrt{\pi/2}, \\ \mathrm{Var}(X) &= \frac{4-\pi}{2}\sigma \\ med &= \sigma\sqrt{\ln(4)}. \end{aligned} \tag{A.4}$$

En s'appuyant sur la fonction de répartition $\mathbb{F}$ associée à $p_{\mathcal{R}}$:

$$\mathbb{F}(x) = 1 - e^{-x^2/(2\sigma^2)} \tag{A.5}$$

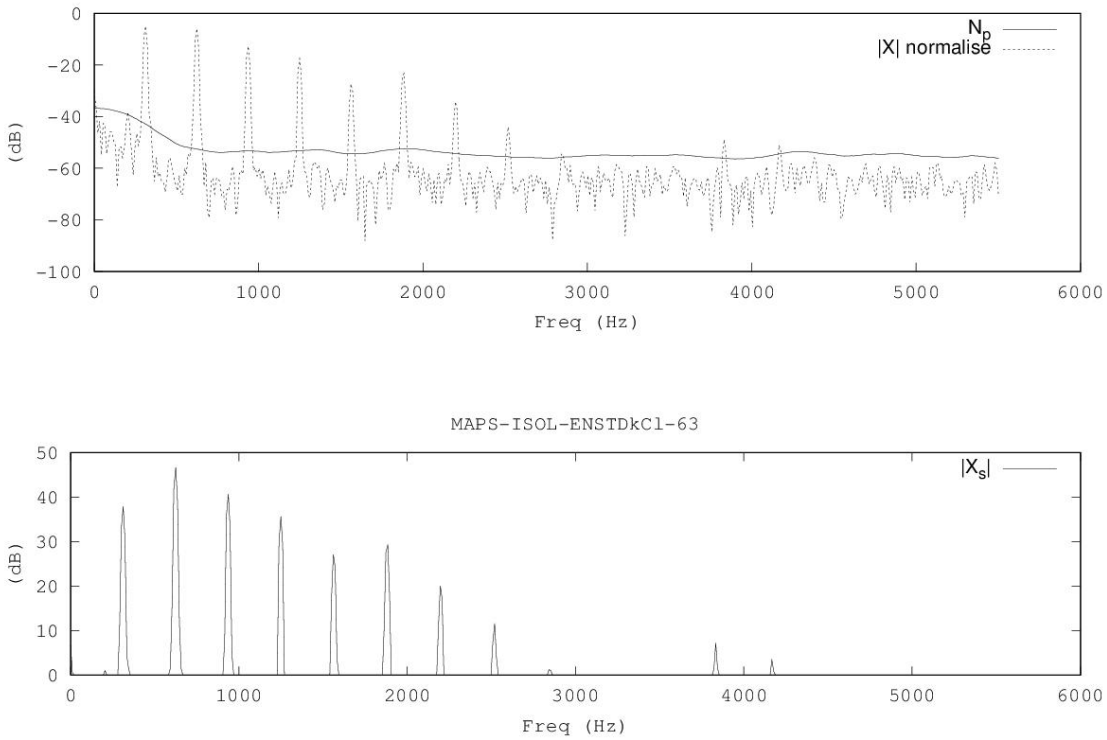


FIGURE A.1 – Spectre normalisé initial et niveau de bruit estimé (figure du haut) et résultats du filtrage médian $|\tilde{X}_s|$(figure du bas) pour une partie de sustain d'un MI_{b4} (midi 63) d'un piano DISKLAVIER tiré de la base *MAPS*, échantillonée à $44100Hz$, calculée avec une fenêtre de Hann de 2^{12} points et tronqué à $F_e/8$, pour $\nu = 1$, $p = 0,999$, et une largeur de bande pour le calcul de la médiane de $300hz$.

on obtient le niveau de filtrage de bruit $\mathcal{N}_p$ associée au p-ième "percentile" :

$$\mathcal{N}_p = \mathbb{F}^{-1}(p) = \sigma\sqrt{-2\log(1-p)}\,; \quad p \in [0,1]. \tag{A.6}$$

En imposant

$$|\tilde{X}_s(f)| = \left\{ \begin{array}{lll} |X(f)| & \forall f \mid & |X(f)| \geq \mathcal{N}_p(f) \\ \mathcal{N}_p(f) & \forall f \mid & |X(f)| < \mathcal{N}_p(f) \end{array} \right. \tag{A.7}$$

on obtient une approximation de la partie déterministe de $|X(f)|$.

Annexe B

Règles de mise à jour

B.1 NMF harmonique

Nous détaillons ici le calcul des règles de mise à jour pour la méthode *NMF harmonique*, telle que décrite §1.1.3. Le seul parmètre est la fréquence fondamentale de chaque atome ($\theta_r = F_{0,r}$ avec $r = [1, ..., 88]$). Les algorithmes multiplicatifs sont décrits §1.1.1.

B.1.1 Minimisation sur Θ

En considérant chaque point *temps-fréquence* du spectrogramme indépendant des autres, on exprime la dérivée de la fonction de coût en $r = r_0$:

$$\left.\frac{\partial \mathcal{C}}{\partial F_{0,r}}\right|_{r_0} = \sum_{ft} \left.\frac{\partial \tilde{V}_{ft}}{\partial F_{0,r}}\right|_{r_0} \tilde{V}_{ft}^{\beta-2}(\tilde{V}_{ft} - V_{ft}). \tag{B.1}$$

Or

$$\begin{aligned} \left.\frac{\partial \tilde{V}_{ft}}{\partial F_{0,r}}\right|_{r_0} &= \left.\frac{\partial W_{fr}^{\theta_r}}{\partial F_{0,r}} H_{rt}\right|_{r_0} \\ &= -H_{r_0 t} \sum_{k=1}^{K_{r_0}} k\, a_k\, g'(f - k\, F_{0,r_0}), \end{aligned} \tag{B.2}$$

d'où

$$\left.\frac{\partial \mathcal{C}}{\partial F_{0,r}}\right|_{r_0} = -\sum_{ft} H_{r_0 t} \sum_{k=1}^{K_{r_0}} k\, a_k\, g'(f - k\, F_{0,r_0}) \tilde{V}_{ft}^{\beta-2}(\tilde{V}_{ft} - V_{ft}) \tag{B.3}$$

Les lobes secondaires de g sont négligés (*ie.* $g'(f) \to g'_\Lambda(f) = g'(f)\mathbb{1}_\Lambda(f)$, avec $\Lambda = [-\frac{T}{2}, \frac{T}{2}]$ le support du lobe principal de g). Sous cette forme, $g'_\Lambda(f)$ prend des valeurs négatives $\forall f > 0$. On pose donc

$$g'_\Lambda(f) = -f\, P(f); \quad P(f) = -\frac{g'_\Lambda(f)}{f} \tag{B.4}$$

Remarque : pour une fenêtre de *Hann*, l'expression de $g'(f)$ est

$$g'(f) = \frac{1}{4\pi^2} \frac{\pi T f \sin(2\pi T f)(T^2 f^2 - 1) - 2\sin^2(\pi T f)(3T^2 f 2 - 1)}{[f(T^2 f^2 - 1)]^3} \tag{B.5}$$

Alors

$$\begin{aligned}\frac{\partial \mathcal{C}}{\partial F_{0,r_0}} &= \sum_{ft} H_{r_0 t} \sum_{k=1}^{K_{r_0}} k\, a_k\, (f - k\, F_{0,r_0}) P(f - k\, F_{0,r_0}) \tilde{V}_{ft}^{\beta-2} (\tilde{V}_{ft} - V_{ft}) \\ &= \mathcal{F}_{r_0+} - \mathcal{F}_{r_0-} \end{aligned} \tag{B.6}$$

avec

$$\begin{aligned}\mathcal{F}_{r_0+} &= \sum_{ft} \sum_{k=1}^{K_{r_0}} H_{r_0 t}\, k\, a_k\, P(f - k\, F_{0,r_0}) \tilde{V}_{ft}^{\beta-2} (f \tilde{V}_{ft} + k\, F_{0,r_0} V_{ft}) \\ \mathcal{F}_{r_0-} &= \sum_{ft} \sum_{k=1}^{K_{r_0}} H_{r_0 t}\, k\, a_k\, P(f - k\, F_{0,r_0}) \tilde{V}_{ft}^{\beta-2} (k\, F_{0,r_0}\, \tilde{V}_{ft} + f\, V_{ft}), \end{aligned} \tag{B.7}$$

ce qui amène à la règle de mise à jour pour F_{0,r_0}

$$F_{0,r_0} \longleftarrow F_{0,r_0} \frac{\mathcal{F}_{r_0-}}{\mathcal{F}_{r_0+}}. \tag{B.8}$$

B.1.2 Minimisation sur H

On se fixe à $r = r_0,\ t = t_0$.

$$\begin{aligned}\frac{\partial \tilde{V}_{ft}}{\partial H_{r_0 t_0}} &= \frac{\partial}{\partial H_{r_0 t_0}} [W_{fr_0}^{\theta_{r_0}} H_{r_0 t_0}] \\ &= W_{fr_0}^{\theta_{r_0}} \end{aligned} \tag{B.9}$$

donc

$$\begin{aligned}\frac{\partial \mathcal{C}}{\partial H_{r_0 t_0}} &= \sum_{f} W_{fr_0}^{\theta_{r_0}} \tilde{V}_{ft_0}^{\beta-2} (\tilde{V}_{ft_0} - V_{ft_0}) \\ &= \mathcal{H}_{r_0 t_0 +} - \mathcal{H}_{r_0 t_0 -} \end{aligned} \tag{B.10}$$

avec

$$\begin{aligned}\mathcal{H}_{r_0 t_0 +} &= \sum_{f} W_{fr_0}^{\theta_{r_0}} \tilde{V}_{ft_0}^{\beta-1} \\ \mathcal{H}_{r_0 t_0 -} &= \sum_{f} W_{fr_0}^{\theta_{r_0}} \tilde{V}_{ft_0}^{\beta-2} V_{ft_0}, \end{aligned} \tag{B.11}$$

ce qui amène à la règle de mise à jour pour $H_{r_0 t_0}$

$$H_{r_0 t_0} \leftarrow H_{r_0 t_0} \frac{\mathcal{H}_{r_0 t_0 -}}{\mathcal{H}_{r_0 t_0 +}}. \tag{B.12}$$

B.1.3 Minimisation sur A

On se fixe à $k = k_0$.

$$\begin{aligned}\frac{\partial \tilde{V}_{ft}}{\partial a_{k_0}} &= \frac{\partial}{\partial a_{k_0}}\left[\sum_{r=1}^{R} H_{rt}\, a_{k_0}\, g(f - k_0\, f_{0,r})\mathbb{1}_{[1,K_r]}(k_0)\right] \\ &= \sum_{r=1}^{R_k} H_{rt}\, g(f - k_0\, F_{0,r}) \end{aligned} \tag{B.13}$$

avec R_k la valeur maximale de r qui assure que l'on respecte (1.12) :

$$R_k = \{\max(r) \mid k_0 \in [1, K_r]\}. \tag{B.14}$$

soit

$$R_k = 1 + 12\log_2(\frac{F_{max}}{k\, F_{ref}}) \tag{B.15}$$

Alors

$$\begin{aligned}\frac{\partial \mathcal{C}}{\partial a_{k_0}} &= \sum_{ft}\sum_{r=1}^{R_k} H_{rt}\, g(f - k_0\, F_{0,r})\tilde{V}_{ft}^{\beta-2}(\tilde{V}_{ft} - V_{ft}) \\ &= \mathcal{A}_{k_0+} - \mathcal{A}_{k_0-} \end{aligned} \tag{B.16}$$

avec

$$\begin{aligned}\mathcal{A}_{k_0+} &= \sum_{ft}\sum_{r=1}^{R_k} H_{rt}\, g(f - k_0\, F_{0,r})\tilde{V}_{ft}^{\beta-1} \\ \mathcal{A}_{k_0-} &= \sum_{ft}\sum_{r=1}^{R_k} H_{rt}\, g(f - k_0\, F_{0,r})\tilde{V}_{ft}^{\beta-2}V_{ft}, \end{aligned} \tag{B.17}$$

ce qui amène à la règle de mise à jour pour a_{k_0}

$$a_{k_0} \leftarrow a_{k_0}\frac{\mathcal{A}_{k_0-}}{\mathcal{A}_{k_0+}}. \tag{B.18}$$

B.1.4 Considérations algorithmiques

On se propose dans cette section de détailler les règles de mise à jour pour l'ensemble des coefficients, d'un point de vue algorithmique, en tirant partie de la puissance de traitement des données matricielles de Matlab. On définit au préalable les matrices suivantes :

$$\mathbf{M}_{ft} = \begin{bmatrix} f(1) & \cdots & f(1) \\ \vdots & \ddots & \vdots \\ f(F) & \cdots & f(F) \end{bmatrix}_{F,T} \tag{B.19}$$

$$\mathbf{M}_{rk} = \begin{bmatrix} 1 & & \cdots & \cdots & & K_1 \\ \vdots & & & & & \vdots \\ 1 & & \cdots & & K_r & 0 \\ \vdots & & & & & \vdots \\ 1 & \cdots & K_R & 0 & \cdots & 0 \end{bmatrix}_{R,K} \tag{B.20}$$

$$\mathbf{M}_{fk}(r_0) = \begin{bmatrix} f(1) - F_{0,r_0} & \cdots & f(1) - K_{r_0}F_{0,r_0} & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ f(F) - F_{0,r_0} & \cdots & f(F) - K_{r_0}F_{0,r_0} & 0 & 0 \end{bmatrix}_{F,K} \quad \forall r_0 \tag{B.21}$$

$$\mathbf{M}_{fr}(k_0) = \begin{bmatrix} f(1) - k_0F_{0,1} & \cdots & f(1) - k_0F_{0,R} \\ \vdots & \ddots & \vdots \\ f(F) - k_0F_{0,1} & \cdots & f(F) - k_0F_{0,R} \end{bmatrix}_{F,R} \quad \forall k_0 \tag{B.22}$$

$$\mathbf{R}_k = \begin{bmatrix} R_1 \\ \vdots \\ R_K \end{bmatrix}_K \tag{B.23}$$

$$\mathbf{V}_{R_k}(k_0) = \begin{bmatrix} \text{is } 1 \leq \mathbf{R}_{k_0} \\ \vdots \\ \text{is } R \leq \mathbf{R}_{k_0} \end{bmatrix}_R \quad \forall k_0 \tag{B.24}$$

$$\begin{aligned} \mathcal{F}_{r_0+} &\longleftarrow {}^t\mathbf{M}_{r_0k} \otimes \mathbf{A} \otimes \left({}^tP(\mathbf{M}_{fk})(\mathbf{M}_{ft} \otimes \tilde{\mathbf{V}}^{\cdot(\beta-1)}){}^t\mathbf{H}_{r_0} + \mathbf{F}_{0,r_0}{}^t\mathbf{M}_{r_0k} \otimes \left({}^tP(\mathbf{M}_{fk})(\tilde{\mathbf{V}}^{\cdot(\beta-2)} \otimes \mathbf{V}){}^t\mathbf{H}_{r_0}\right)\right) \\ \mathcal{F}_{r_0-} &\longleftarrow {}^t\mathbf{M}_{r_0k} \otimes \mathbf{A} \otimes \left({}^tP(\mathbf{M}_{fk})(\mathbf{M}_{ft} \otimes \tilde{\mathbf{V}}^{\cdot(\beta-2)} \otimes \mathbf{V}){}^t\mathbf{H}_{r_0} + \mathbf{F}_{0,r_0}{}^t\mathbf{M}_{r_0k} \otimes \left({}^tP(\mathbf{M}_{fk})(\tilde{\mathbf{V}}^{\cdot(\beta-1)}){}^t\mathbf{H}_{r_0}\right)\right) \\ \mathbf{F}_{0,r_0} &\longleftarrow \mathbf{F}_{0,r_0} \otimes \frac{\mathcal{F}_{r_0-}}{\mathcal{F}_{r_0+}}. \end{aligned} \tag{B.25}$$

$$\mathbf{H} \longleftarrow \mathbf{H} \otimes \frac{{}^t\mathbf{W}\tilde{\mathbf{V}}^{\cdot(\beta-1)}}{{}^t\mathbf{W}(\mathbf{V} \otimes \tilde{\mathbf{V}}^{\cdot(\beta-2)})}. \tag{B.26}$$

$$\mathbf{M}_{VR_k}(k_0) = \begin{bmatrix} \mathbf{V}_{R_{k_0}}(1) & \cdots & \mathbf{V}_{R_{k_0}}(1) \\ \vdots & \vdots & \vdots \\ \mathbf{V}_{R_{k_0}}(R) & \cdots & \mathbf{V}_{R_{k_0}}(R) \end{bmatrix}_{R,T} \quad \forall k_0 \tag{B.27}$$

$$\mathbf{A}_{k0} \longleftarrow \mathbf{A}_{k0} \otimes \frac{\sum_{t=1}^{T}\sum_{f=1}^{F}\left[(g(\mathbf{M}_{fr})(\mathbf{H} \otimes \mathbf{M}_{VR_k})) \otimes \mathbf{V} \otimes \tilde{\mathbf{V}}^{\cdot(\beta-2)}\right]}{\sum_{t=1}^{T}\sum_{f=1}^{F}\left[(g(\mathbf{M}_{fr})(\mathbf{H} \otimes \mathbf{M}_{VR_k})) \otimes \tilde{\mathbf{V}}^{\cdot(\beta-1)}\right]}. \tag{B.28}$$

end for

B.2 NMF sous contrainte stricte d'inharmonicité

B.2.1 Minimisation sur Θ

Minimisation sur $F_{0,r}$

On fixe $r = r_0$:

$$\frac{\partial \mathcal{C}}{\partial F_{0,r_0}} = \sum_{ft} \frac{\partial \tilde{V}_{ft}}{\partial F_{0,r_0}} \tilde{V}_{ft}^{\beta-2}(\tilde{V}_{ft} - V_{ft}) \tag{B.29}$$

$$\begin{aligned}\frac{\partial \tilde{V}_{ft}}{\partial F_{0,r_0}} &= \frac{\partial W_{fr_0}^{\theta_{r_0}}}{\partial F_{0,r_0}} H_{r_0 t} \\ &= -H_{r_0 t} \sum_{k=1}^{K_{r_0}} k \sqrt{1 + B_{r_0} k^2}\, a_k\, g'(f - f_{r_0,k}) \end{aligned} \tag{B.30}$$

Les lobes secondaires de g sont négligés (*ie.* $g'(f) \to g'_\Lambda(f) = g'(f)\mathbb{1}_\Lambda(f)$, avec $\Lambda = [-\frac{T}{2}, \frac{T}{2}]$ le support du lobe principal de g). Sous cette forme, $g'_\Lambda(f)$ prend des valeurs négatives $\forall f > 0$. On pose donc

$$g'_\Lambda(f) = -f\, P(f); \quad P(f) = -\frac{g'_\Lambda(f)}{f} \tag{B.31}$$

Remarque : pour une fenêtre de *Hann*, l'expression de $g'(f)$ est

$$g'(f) = \frac{1}{4\pi^2} \frac{\pi T f \sin(2\pi T f)(T^2 f^2 - 1) - 2\sin^2(\pi T f)(3T^2 f 2 - 1)}{[f(T^2 f^2 - 1)]^3} \tag{B.32}$$

Alors

$$\begin{aligned}\frac{\partial \mathcal{C}}{\partial \theta_{r_0}} &= \sum_{ft} H_{r_0 t} \sum_{k=1}^{K_{r_0}} k \sqrt{1 + B_{r_0} k^2}\, (f - f_{r_0,k})\, a_k\, P(f - f_{r_0,k}) \tilde{V}_{ft}^{\beta-2} (\tilde{V}_{ft} - V_{ft}) \\ &= \mathcal{G}_{r_0} - \mathcal{F}_{r_0} \end{aligned} \tag{B.33}$$

avec

$$\begin{aligned}\mathcal{G}_{r_0} &= \sum_{ft} \sum_{k=1}^{K_{r_0}} H_{r_0 t}\, k \sqrt{1 + B_{r_0} k^2}\, a_k\, P(f - f_{r_0,k}) \tilde{V}_{ft}^{\beta-2} (f \tilde{V}_{ft} + f_{r_0,k} V_{ft}) \\ \mathcal{F}_{r_0} &= \sum_{ft} \sum_{k=1}^{K_{r_0}} H_{r_0 t}\, k \sqrt{1 + B_{r_0} k^2}\, a_k\, P(f - f_{r_0,k}) \tilde{V}_{ft}^{\beta-2} (\, f_{r_0,k}\, \tilde{V}_{ft} + f\, V_{ft}), \end{aligned} \tag{B.34}$$

ce qui amène à la règle de mise à jour pour F_{0,r_0}

$$F_{0,r_0} \leftarrow F_{0,r_0} \frac{\mathcal{F}_{r_0}}{\mathcal{G}_{r_0}}. \tag{B.35}$$

Minimisation sur B_r

On fixe $r = r_0$:

$$\frac{\partial \mathcal{C}}{\partial B_{r_0}} = \sum_{ft} \frac{\partial \tilde{V}_{ft}}{\partial B_{r_0}} \tilde{V}_{ft}^{\beta-2} (\tilde{V}_{ft} - V_{ft}) \tag{B.36}$$

$$\begin{aligned}\frac{\partial \tilde{V}_{ft}}{\partial B_{r_0}} &= \frac{\partial W_{fr_0}^{\theta_{r_0}}}{\partial B_{r_0}} H_{r_0 t} \\ &= -\frac{1}{2} H_{r_0 t} \sum_{k=1}^{K_{r_0}} k^3 F_{0,r_0}\, a_k\, \frac{g'(f - f_{r_0,k})}{\sqrt{1 + B_{r_0} k^2}} \end{aligned} \tag{B.37}$$

Alors

$$\begin{aligned}\frac{\partial \mathcal{C}}{\partial \theta_{r_0}} &= \frac{1}{2}\sum_{ft} H_{r_0 t} \sum_{k=1}^{K_{r_0}} k^3 F_{0,r_0}\, a_k\, \frac{P(f-f_{r_0,k})}{\sqrt{1+B_{r_0}k^2}}\,(f-f_{r_0,k})\tilde{V}_{ft}^{\beta-2}(\tilde{V}_{ft}-V_{ft}) \\ &= \mathcal{G}_{r_0} - \mathcal{F}_{r_0} \end{aligned} \tag{B.38}$$

avec

$$\begin{aligned}\mathcal{G}_{r_0} &= \frac{1}{2}\sum_{ft}\sum_{k=1}^{K_{r_0}} H_{r_0 t}\, k^3 F_{0,r_0}\, a_k\, \frac{P(f-f_{r_0,k})}{\sqrt{1+B_{r_0}k^2}}\,\tilde{V}_{ft}^{\beta-2}(f\tilde{V}_{ft}+f_{r_0,k}V_{ft}) \\ \mathcal{F}_{r_0} &= \frac{1}{2}\sum_{ft}\sum_{k=1}^{K_{r_0}} H_{r_0 t}\, k^3 F_{0,r_0}\, a_k\, \frac{P(f-f_{r_0,k})}{\sqrt{1+B_{r_0}k^2}}\,\tilde{V}_{ft}^{\beta-2}(\, f_{r_0,k}\,\tilde{V}_{ft}+f\,V_{ft}), \end{aligned} \tag{B.39}$$

ce qui amène à la règle de mise à jour pour F_{0,r_0}

$$B_{r_0} \leftarrow B_{r_0}\frac{\mathcal{F}_{r_0}}{\mathcal{G}_{r_0}}. \tag{B.40}$$

B.2.2 Minimisation sur H

On se fixe à $H_{r_0 t_0}$.

$$\begin{aligned}\frac{\partial \tilde{V}_{ft}}{\partial H_{r_0 t_0}} &= \frac{\partial}{\partial H_{r_0 t_0}}[W_{fr_0}^{\theta_{r_0}} H_{r_0 t_0}] \\ &= W_{fr_0}^{\theta_{r_0}} \end{aligned} \tag{B.41}$$

donc

$$\begin{aligned}\frac{\partial \mathcal{C}}{\partial H_{r_0 t_0}} &= \sum_f W_{fr_0}^{\theta_{r_0}} \tilde{V}_{ft_0}^{\beta-2}(\tilde{V}_{ft_0}-V_{ft_0}) \\ &= \mathcal{P}r_0 - \mathcal{M}_{r_0} \end{aligned} \tag{B.42}$$

avec

$$\begin{aligned}\mathcal{P}_{r_0 t_0} &= \sum_f W_{fr_0}^{\theta_{r_0}} \tilde{V}_{ft_0}^{\beta-1} \\ \mathcal{M}_{r_0 t_0} &= \sum_f W_{fr_0}^{\theta_{r_0}} \tilde{V}_{ft_0}^{\beta-2} V_{ft_0}, \end{aligned} \tag{B.43}$$

ce qui amène à la règle de mise à jour pour $H_{r_0 t_0}$

$$H_{r_0 t_0} \leftarrow H_{r_0 t_0}\frac{\mathcal{M}_{r_0 t_0}}{\mathcal{P}_{r_0 t_0}}. \tag{B.44}$$

B.2.3 Minimisation sur A

On se fixe à $k = k_0$.

$$\begin{aligned}\frac{\partial \tilde{V}_{ft}}{\partial a_{k_0}} &= \frac{\partial}{\partial a_{k_0}}\left[\sum_{r=1}^{R} H_{rt}\, a_{k_0}\, g(f - k_0\, f_{r,k_0}) \mathbb{1}_{[1,K_r]}(k_0)\right] \\ &= \sum_{r=1}^{R_k} H_{rt}\, g(f - k_0\, f_{r,k_0}) \end{aligned} \tag{B.45}$$

avec R_k la valeur maximale de r qui assure que l'on respecte (1.12) :

$$R_k = \{\max(r) \mid k_0 \in [1, K_r]\}. \tag{B.46}$$

soit

$$R_k = 1 + 12 \log_2(\frac{F_{max}}{k\, F_{ref}}) \tag{B.47}$$

Alors

$$\begin{aligned}\frac{\partial \mathcal{C}}{\partial a_{k_0}} &= \sum_{ft} \sum_{r=1}^{R_k} H_{rt}\, g(f - k_0\, f_{r,k_0}) \tilde{V}_{ft}^{\beta-2} (\tilde{V}_{ft} - V_{ft}) \\ &= \mathcal{A}_{k_0+} - \mathcal{A}_{k_0-} \end{aligned} \tag{B.48}$$

avec

$$\begin{aligned}\mathcal{A}_{k_0+} &= \sum_{ft} \sum_{r=1}^{R_k} H_{rt}\, g(f - k_0\, f_{r,k_0}) \tilde{V}_{ft}^{\beta-1} \\ \mathcal{A}_{k_0-} &= \sum_{ft} \sum_{r=1}^{R_k} H_{rt}\, g(f - k_0\, f_{r,k_0}) \tilde{V}_{ft}^{\beta-2} V_{ft}, \end{aligned} \tag{B.49}$$

ce qui amène à la règle de mise à jour pour a_{k_0}

$$a_{k_0} \leftarrow a_{k_0} \frac{\mathcal{A}_{k_0-}}{\mathcal{A}_{k_0+}}. \tag{B.50}$$

B.2.4 Considérations algorithmiques

$$\mathbf{M}_{ft} = \begin{bmatrix} f(1) & \cdots & f(1) \\ \vdots & \ddots & \vdots \\ f(F) & \cdots & f(F) \end{bmatrix}_{F,T} \tag{B.51}$$

$$\mathbf{M}_{rk} = \begin{bmatrix} 1 & & \cdots & \cdots & & K_1 \\ \vdots & & & & & \vdots \\ 1 & & \cdots & & K_r & 0 \\ \vdots & & & & & \vdots \\ 1 & \cdots & K_R & 0 & \cdots & 0 \end{bmatrix}_{R,K} \tag{B.52}$$

$$\mathbf{R}_k = \begin{bmatrix} R_1 \\ \vdots \\ R_K \end{bmatrix}_K \tag{B.53}$$

$$\mathbf{M}_{fk}(r_0) = \begin{bmatrix} f(1) - f_{r_0,1} & \cdots & f(1) - f_{r_0,K_{r_0}} & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ f(F) - f_{r_0,1} & \cdots & f(F) - f_{r_0,K_{r_0}} & 0 & 0 \end{bmatrix}_{F,K} \quad \forall r_0 \tag{B.54}$$

$$\mathbf{M}_{fr}(k_0) = \begin{bmatrix} f(1) - f_{1,k_0} & \cdots & f(1) - f_{R,k_0} \\ \vdots & \ddots & \vdots \\ f(F) - f_{1,k_0} & \cdots & f(F) - f_{R,k_0} \end{bmatrix}_{F,R} \quad \forall k_0 \tag{B.55}$$

$$\mathbf{V}_{R_k}(k_0) = \begin{bmatrix} \text{is } 1 \leq \mathbf{R}_{k_0} \\ \vdots \\ \text{is } R \leq \mathbf{R}_{k_0} \end{bmatrix}_R \quad \forall k_0 \tag{B.56}$$

H

$$\mathbf{H} \longleftarrow \mathbf{H} \otimes \frac{{}^t\mathbf{W}(\mathbf{V} \otimes \tilde{\mathbf{V}}^{\cdot(\beta-2)})}{{}^t\mathbf{W}(\tilde{\mathbf{V}}^{\cdot(\beta-1)})} \tag{B.57}$$

$a_{r,k}$

for $k_0 = 1 \to K$

$$\mathbf{M}_{VR_k}(k_0) = \begin{bmatrix} \mathbf{V}_{R_{k_0}}(1) & \cdots & \mathbf{V}_{R_{k_0}}(1) \\ \vdots & \vdots & \vdots \\ \mathbf{V}_{R_{k_0}}(R) & \cdots & \mathbf{V}_{R_{k_0}}(R) \end{bmatrix}_{R,T} \quad \forall k_0 \tag{B.58}$$

$$\mathbf{A}_{k0} \longleftarrow \mathbf{A}_{k0} \otimes \frac{\sum_{t=1}^{T} \sum_{f=1}^{F} \left[(g(\mathbf{M}_{fr})(\mathbf{H} \otimes \mathbf{M}_{VR_k})) \otimes \mathbf{V} \otimes \tilde{\mathbf{V}}^{\cdot(\beta-2)} \right]}{\sum_{t=1}^{T} \sum_{f=1}^{F} \left[(g(\mathbf{M}_{fr})(\mathbf{H} \otimes \mathbf{M}_{VR_k})) \otimes \tilde{\mathbf{V}}^{\cdot(\beta-1)} \right]}. \tag{B.59}$$

end for

$F_{0,r}$

for $r_0 = 1 \to R$

Mise à jour de $\mathbf{M}_{fk}(r_0)$

$$\begin{aligned}
\mathcal{F}_{0-} &= \sum_{\forall k}\Big(\mathbf{M}_{r_0,k}\otimes\sqrt{1+B_{r_0}\mathbf{M}_{r_0,k}^{.2}}\otimes a_k\otimes\Big[{\mathbf{H}_{r_0}}^t\Big(\mathbf{M}_{ft}\otimes\mathbf{V}\otimes\tilde{\mathbf{V}}^{.(\beta-2)}\Big)P(\mathbf{M}_{fk})+\\
&\quad +f_{r_0,k}\otimes\Big({\mathbf{H}_{r_0}}^t\Big(\tilde{\mathbf{V}}^{.(\beta-1)}\Big)P(\mathbf{M}_{fk})\Big)\Big]\Big)\\
\mathcal{F}_{0+} &= \sum_{\forall k}\Big(\mathbf{M}_{r_0,k}\otimes\sqrt{1+B_{r_0}\mathbf{M}_{r_0,k}^{.2}}\otimes a_k\otimes\Big[{\mathbf{H}_{r_0}}^t\Big(\mathbf{M}_{ft}\otimes\tilde{\mathbf{V}}^{.(\beta-1)}\Big)P(\mathbf{M}_{fk})+\\
&\quad +f_{r_0,k}\otimes\Big({\mathbf{H}_{r_0}}^t\Big(\mathbf{V}\otimes\tilde{\mathbf{V}}^{.(\beta-2)}\Big)P(\mathbf{M}_{fk})\Big)\Big]\Big)\\
F_{0,r_0} &\longleftarrow F_{0,r_0}\otimes\frac{\mathcal{F}_{0-}}{\mathcal{F}_{0+}} \qquad \text{(B.60)}
\end{aligned}$$

end for

B_r

for $r_0 = 1 \to R$

Mise à jour de $\mathbf{M}_{fk}(r_0)$

$$\begin{aligned}
\mathcal{B}_{0-} &= \sum_{\forall k}\Big(\mathbf{M}_{r_0,k}^{.3}\otimes(1+B_{r_0}\mathbf{M}_{r_0,k}^{.2})^{.-\frac{1}{2}}\otimes a_k\otimes\Big[{\mathbf{H}_{r_0}}^t\Big(\mathbf{M}_{ft}\otimes\mathbf{V}\otimes\tilde{\mathbf{V}}^{.(\beta-2)}\Big)P(\mathbf{M}_{fk})+\\
&\quad +{}_{r_0,k}\otimes\Big({\mathbf{H}_{r_0}}^t\Big(\tilde{\mathbf{V}}^{.(\beta-1)}\Big)P(\mathbf{M}_{fk})\Big)\Big]\\
\mathcal{B}_{0+} &= \sum_{\forall k}\Big(\mathbf{M}_{r_0,k}^{.3}\otimes(1+B_{r_0}\mathbf{M}_{r_0,k}^{.2})^{.-\frac{1}{2}}\otimes a_k\otimes\Big[{\mathbf{H}_{r_0}}^t\Big(\mathbf{M}_{ft}\otimes\tilde{\mathbf{V}}^{.(\beta-1)}\Big)P(\mathbf{M}_{fk})+\\
&\quad +f_{r_0,k}\otimes\Big({\mathbf{H}_{r_0}}^t\Big(\mathbf{V}\otimes\tilde{\mathbf{V}}^{.(\beta-2)}\Big)P(\mathbf{M}_{fk})\Big)\Big]\Big)\\
B_{r_0} &\longleftarrow B_{r_0}\otimes\frac{\mathcal{B}_{0-}}{\mathcal{B}_{0+}} \qquad \text{(B.61)}
\end{aligned}$$

end for

B.3 NMF sous contrainte relaxée d'inharmonicité

B.3.1 Minimisation de $\mathcal{C}_0$

Minimisation sur $a_{r,k}$

On fixe $r = r_0,\ k = k_0$:

$$\begin{aligned}
\frac{\partial\mathcal{C}_0}{\partial a_{r_0,k_0}} &= \frac{1}{FT}\sum_{ft}H_{r_0,t}g(f-f_{r_0,k_0})\tilde{V}_{ft}^{\beta_0-2}(\tilde{V}_{ft}-V_{ft})\\
&= \mathcal{A}_{0+}(r_0,k_0)-\mathcal{A}_{0-}(r_0,k_0) \qquad \text{(B.62)}
\end{aligned}$$

avec

$$\begin{aligned}\mathcal{A}_{0+}(r_0,k_0) &= \frac{1}{FT}\sum_{ft} H_{r_0,t}\; g(f-f_{r_0,k_0})\tilde{V}_{ft}^{\beta_0-1}\\ \mathcal{A}_{0-}(r_0,k_0) &= \frac{1}{FT}\sum_{ft} H_{r_0,t}\; g(f-f_{r_0,k_0})\tilde{V}_{ft}^{\beta_0-2}V_{ft}\end{aligned} \tag{B.63}$$

ce qui amène à la règle de mise à jour pour F_{0,r_0}

$$a_{r_0,k_0} \leftarrow a_{r_0,k_0}\frac{\mathcal{A}_{0-}(r_0,k_0)}{\mathcal{A}_{0+}(r_0,k_0)}. \tag{B.64}$$

Minimisation sur $f_{r,k}$

On fixe $r=r_0,\ k=k_0$:

$$\begin{aligned}\frac{\partial \mathcal{C}_0}{\partial f_{r_0,k_0}} &= -\frac{1}{FT}\sum_{ft} H_{r_0,t}\; a_{r_0,k_0}\; g'(f-f_{r_0,k_0})\tilde{V}_{ft}^{\beta_0-2}(\tilde{V}_{ft}-V_{ft})\\ &= \frac{1}{FT}\sum_{ft} H_{r_0,t}\; a_{r_0,k_0}\; P(f-f_{r_0,k_0})\tilde{V}_{ft}^{\beta_0-2}(\tilde{V}_{ft}-V_{ft})(f-f_{r_0,k_0})\\ &= \mathcal{F}_{0+}(r_0,k_0)-\mathcal{F}_{0-}(r_0,k_0)\end{aligned} \tag{B.65}$$

avec

$$\begin{aligned}\mathcal{F}_{0+}(r_0,k_0) &= \frac{1}{FT}\sum_{ft} H_{r_0,t}\; a_{r_0,k_0}\; P(f-f_{r_0,k_0})\tilde{V}_{ft}^{\beta_0-2}(f\;\tilde{V}_{ft}+f_{r_0,k_0}\;V_{ft})\\ \mathcal{F}_{0-}(r_0,k_0) &= \frac{1}{FT}\sum_{ft} H_{r_0,t}\; a_{r_0,k_0}\; P(f-f_{r_0,k_0})\tilde{V}_{ft}^{\beta_0-2}(f_{r_0,k_0}\;\tilde{V}_{ft}+f\;V_{ft})\end{aligned} \tag{B.66}$$

Minimisation sur H

On se fixe à $H_{r_0t_0}$.

$$\begin{aligned}\frac{\partial \tilde{V}_{ft}}{\partial H_{r_0t_0}} &= \frac{\partial}{\partial H_{r_0t_0}}[W_{fr_0}^{\theta_{r_0}}H_{r_0t_0}]\\ &= W_{fr_0}^{\theta_{r_0}}\end{aligned} \tag{B.67}$$

donc

$$\begin{aligned}\frac{\partial \mathcal{C}_0}{\partial H_{r_0t_0}} &= \frac{1}{FT}\sum_{f} W_{fr_0}^{\theta_{r_0}}\tilde{V}_{ft_0}^{\beta_0-2}(\tilde{V}_{ft_0}-V_{ft_0})\\ &= \mathcal{H}_{0+}(r_0,t_0)-\mathcal{H}_{0-}(r_0,t_0)\end{aligned} \tag{B.68}$$

avec

$$\begin{aligned}\mathcal{H}_{0+}(r_0,t_0) &= \frac{1}{FT}\sum_{f} W_{fr_0}^{\theta_{r_0}}\tilde{V}_{ft_0}^{\beta_0-1}\\ \mathcal{H}_{0-}(r_0,t_0) &= \frac{1}{FT}\sum_{f} W_{fr_0}^{\theta_{r_0}}\tilde{V}_{ft_0}^{\beta_0-2}V_{ft_0},\end{aligned} \tag{B.69}$$

ce qui amène à la règle de mise à jour pour $H_{r_0t_0}$

$$H_{r_0t_0} \leftarrow H_{r_0t_0} \frac{\mathcal{H}_{0-}(r_0,t_0)}{\mathcal{H}_{0+}(r_0,t_0)}. \tag{B.70}$$

B.3.2 Minimisation de $\mathcal{C}_1$

Minimisation sur $f_{r,k}$

On fixe $r = r_0,\ k = k_0$:

$$\begin{aligned} \frac{\partial \mathcal{C}_1}{\partial f_{r_0,k_0}} &= \frac{1}{K_{r_0}(\beta_1-1)}(f_{r_0,k_0}^{(\beta_1-1)} - (k_0 F_{0,r_0}\sqrt{1+B_{r_0}k_0^2})^{(\beta_1-1)}) \\ &= \mathcal{F}_{1+}(r_0,k_0) - \mathcal{F}_{1-}(r_0,k_0) \end{aligned} \tag{B.71}$$

avec

$$\mathcal{F}_{1+}(r_0,k_0) = \frac{f_{r_0,k_0}^{(\beta_1-1)}}{K_{r_0}(\beta_1-1)} \tag{B.72}$$

$$\mathcal{F}_{1-}(r_0,k_0) = \frac{(k_0 F_{0,r_0}\sqrt{1+B_{r_0}k_0^2})^{(\beta_1-1)}}{K_{r_0}(\beta_1-1)} \tag{B.73}$$

ce qui amène à la règle de mise à jour pour f_{r_0,k_0}

$$f_{r_0,k_0} \leftarrow f_{r_0,k_0} \frac{\mathcal{F}_{0-}(r_0,k_0) + \lambda_1 \mathcal{F}_{1-}(r_0,k_0)}{\mathcal{F}_{0+}(r_0,k_0) + \lambda_1 \mathcal{F}_{1+}(r_0,k_0)}. \tag{B.74}$$

Minimisation sur B_r

On fixe $r = r_0$:

$$\frac{\partial \mathcal{C}_1}{\partial B_{r_0}} = \frac{1}{2K_{r_0}} \sum_{k=1}^{K_{r_0}} k^{\beta_1+1} F_{0,r_0}^{\beta_1-1} (1+B_{r_0}k^2)^{\beta_1-\frac{5}{2}} (kF_{0,r_0}\sqrt{1+B_{r_0}k^2} - f_{r_0,k}) \tag{B.75}$$

$$= \mathcal{B}_{1+}(r_0) - \mathcal{B}_{1-}(r_0) \tag{B.76}$$

avec

$$\mathcal{B}_{1+}(r_0,k_0) = \frac{1}{2K_{r_0}} \sum_{k=1}^{K_{r_0}} k^{\beta_1+2} F_{0,r_0}^{\beta_1} (1+B_{r_0}k^2)^{\beta_1-2} \tag{B.77}$$

$$\mathcal{B}_{1-}(r_0,k_0) = \frac{1}{2K_{r_0}} \sum_{k=1}^{K_{r_0}} k^{\beta_1+1} F_{0,r_0}^{\beta_1-1} (1+B_{r_0}k^2)^{\beta_1-\frac{5}{2}} f_{r_0,k} \tag{B.78}$$

ce qui amène à la règle de mise à jour pour B_{r_0}

$$B_{r_0} \leftarrow B_{r_0} \frac{\mathcal{B}_{1-}(r_0)}{\mathcal{B}_{1+}(r_0)}. \tag{B.79}$$

Minimisation sur $F_{0,r}$

On fixe $r = r_0$:

$$\frac{\partial \mathcal{C}_1}{\partial F_{0,r_0}} = \frac{1}{K_{r_0}} \sum_{k=1}^{K_{r_0}} k^{\beta_1 - 1} F_{0,r_0}^{\beta_1 - 2} (1 + B_{r_0} k^2)^{\beta_1 - \frac{3}{2}} (k F_{0,r_0} \sqrt{1 + B_{r_0} k^2} - f_{r_0,k}) \tag{B.80}$$

$$= \mathcal{F}'_{1+}(r_0) - \mathcal{F}'_{1-}(r_0) \tag{B.81}$$

avec

$$\mathcal{F}'_{1+}(r_0) = \frac{1}{K_{r_0}} \sum_{k=1}^{K_{r_0}} k^{\beta_1} F_{0,r_0}^{\beta_1 - 1} (1 + B_{r_0} k^2)^{\beta_1 - 1} \tag{B.82}$$

$$\mathcal{F}'_{1-}(r_0) = \frac{1}{K_{r_0}} \sum_{k=1}^{K_{r_0}} k^{\beta_1 - 1} F_{0,r_0}^{\beta_1 - 2} (1 + B_{r_0} k^2)^{\beta_1 - \frac{3}{2}} f_{r_0,k} \tag{B.83}$$

ce qui amène à la règle de mise à jour pour F_{0,r_0}

$$F_{0,r_0} \leftarrow F_{0,r_0} \frac{\mathcal{F}'_{1-}(r_0)}{\mathcal{F}'_{1+}(r_0)}. \tag{B.84}$$

B.3.3 Considérations algorithmiques

$$\mathbf{M}_{ft} = \begin{bmatrix} f(1) & \cdots & f(1) \\ \vdots & \ddots & \vdots \\ f(F) & \cdots & f(F) \end{bmatrix}_{F,T} \tag{B.85}$$

$$\mathbf{M}_{rk} = \begin{bmatrix} 1 & & \cdots & \cdots & & K_1 \\ \vdots & & & & & \vdots \\ 1 & & \cdots & & K_r & 0 \\ \vdots & & & & & \vdots \\ 1 & \cdots & K_R & 0 & \cdots & 0 \end{bmatrix}_{R,K} \tag{B.86}$$

H

$$H \longleftarrow H \otimes \frac{{}^t W(\tilde{V}^{\cdot(\beta_0 - 1)})}{{}^t W(V \otimes \tilde{V}^{\cdot(\beta_0 - 2)})} \tag{B.87}$$

$a_{r,k}$

for $r_0 = 1 \to R$

$$\mathbf{M}_{fk}(r_0) = \begin{bmatrix} f(1) - f_{r_0,1} & \cdots & f(1) - f_{r_0,K} \\ \vdots & \ddots & \vdots \\ f(F) - f_{r_0,1} & \cdots & f(F) - f_{r_0,K} \end{bmatrix} \tag{B.88}$$

$$a_{r_0,k} \longleftarrow a_{r_0,k} \otimes \frac{H_{r_0}{}^t(V \otimes \tilde{V}^{\cdot(\beta_0-2)})\, g(\mathbf{M}_{fk})}{H_{r_0}{}^t(\tilde{V}^{\cdot(\beta_0-1)})\, g(\mathbf{M}_{fk})} \tag{B.89}$$

end for

$f_{r,k}$

for $r_0 = 1 \to R$

Mise à jour de $\mathbf{M}_{fk}(r_0)$

$$\begin{aligned}
\mathcal{F}_{0-} &= a_{r_0,k} \otimes \left[{}^tP(\mathbf{M}_{fk})(\mathbf{M}_{ft} \otimes V \otimes \tilde{V}^{\cdot(\beta_0-2)}){}^tH_{r_0} + f_{r_0,k} \otimes ({}^tP(\mathbf{M}_{fk})\tilde{V}^{\cdot(\beta_0-1)}{}^tH_{r_0}) \right] \\
\mathcal{F}_{0+} &= a_{r_0,k} \otimes \left[{}^tP(\mathbf{M}_{fk})(\mathbf{M}_{ft} \otimes \tilde{V}^{\cdot(\beta_0-1)}){}^tH_{r_0} + f_{r_0,k} \otimes ({}^tP(\mathbf{M}_{fk})(V \otimes \tilde{V}^{\cdot(\beta_0-2)}){}^tH_{r_0}) \right] \\
\mathcal{F}_{1-} &= \left(F_{0,r_0}\mathbf{M}_{r_0,k} \otimes \sqrt{1 + B_{r_0}\mathbf{M}_{r_0,k}^{\cdot 2}} \right)^{\cdot(\beta_1-1)} \\
\mathcal{F}_{1-} &= f_{r_0,k}^{\cdot(\beta_1-1)} \\
f_{r_0,k} &\longleftarrow f_{r_0,k} \otimes \frac{\mathcal{F}_{0-} + \lambda_1\mathcal{F}_{1-}}{\mathcal{F}_{0+} + \lambda_1\mathcal{F}_{1+}}
\end{aligned} \tag{B.90}$$

end for

B_r

$$\mathbf{M}_{bk} = \begin{bmatrix} B_{r_1} & \cdots & B_{r_1} \\ \vdots & \ddots & \vdots \\ B_{r_R} & \cdots & B_{r_R} \end{bmatrix}_{R,K} \tag{B.91}$$

et

$$\mathbf{M}_{f_0k} = \begin{bmatrix} F_{0,r_1} & \cdots & F_{0,r_1} \\ \vdots & \ddots & \vdots \\ F_{0,r_R} & \cdots & F_{0,r_R} \end{bmatrix}_{R,K} \tag{B.92}$$

$$B_r \longleftarrow B_r \otimes \frac{\sum_{\forall\, k} \mathbf{M}_{rk}^{\cdot(\beta_1+1)} \otimes \mathbf{M}_{f_0k}^{\cdot(\beta_1-1)} \otimes \left(1 + \mathbf{M}_{bk} \otimes \mathbf{M}_{rk}^{\cdot 2}\right)^{\cdot(\beta_1-\frac{5}{2})} \otimes f_{r,k}}{\sum_{\forall\, k} \mathbf{M}_{rk}^{\cdot(\beta_1+2)} \otimes \mathbf{M}_{f_0k}^{\cdot(\beta_1)} \otimes \left(1 + \mathbf{M}_{bk} \otimes \mathbf{M}_{rk}^{\cdot 2}\right)^{\cdot(\beta_1-2)}} \tag{B.93}$$

$F_{0,r}$

$$F_{0,r} \longleftarrow F_{0,r} \otimes \frac{\sum_{\forall\, k} \mathbf{M}_{rk}^{\cdot(\beta_1-1)} \otimes \mathbf{M}_{f_0k}^{\cdot(\beta_1-2)} \otimes \left(1 + \mathbf{M}_{bk} \otimes \mathbf{M}_{rk}^{\cdot 2}\right)^{\cdot(\beta_1-\frac{3}{2})} \otimes f_{r,k}}{\sum_{\forall\, k} \mathbf{M}_{rk}^{\cdot\beta_1} \otimes \mathbf{M}_{f_0k}^{\cdot(\beta_1-1)} \otimes \left(1 + \mathbf{M}_{bk} \otimes \mathbf{M}_{rk}^{\cdot 2}\right)^{\cdot(\beta_1-1)}} \tag{B.94}$$

B.4 Paramétrisation de H sous forme d'exponentielles décroissantes

A_{r_0,i_0}

$$\frac{\partial \mathcal{C}_0}{\partial A_{r_0,i_0}} = \frac{1}{FT} \sum_{ft} W_{fr_0}^{\theta_{r_0}} e^{-\delta_{r_0}(t-O_{i_0})} \tilde{V}_{ft}^{\beta_0-2} (\tilde{V}_{ft} - V_{ft}) \tag{B.95}$$

δ_{r_0}

$$\frac{\partial \mathcal{C}_0}{\partial \delta_{r_0}} = \frac{1}{FT} \sum_{ft} W_{fr_0}^{\theta_{r_0}} \sum_{i=1}^{I_t} A_{r_0,i}(t - O_i) e^{-\delta_{r_0}(t-O_i)} \tilde{V}_{ft}^{\beta_0-2} (V_{ft} - \tilde{V}_{ft}) \tag{B.96}$$

B.4.1 Algorithmie

$$\begin{aligned}
\mathbf{V}_{ti} &= \begin{bmatrix} \max(i) \mid O_i < t(1) \\ \vdots \\ \max(i) \mid O_i < t(T) \end{bmatrix}_{T,1} \\
\mathbf{M}_{ti} &= \begin{bmatrix} 1 < \mathbf{V}_{1i} & \cdots & I < \mathbf{V}_{1i} \\ \vdots & & \vdots \\ 1 < \mathbf{V}_{Ti} & \cdots & I < \mathbf{V}_{Ti} \end{bmatrix}_{T,I} \\
\mathbf{M}_{to} &= \begin{bmatrix} t(1) - O_1 & \cdots & t(1) - O_I \\ \vdots & & \vdots \\ t(T) - O_1 & \cdots & t(T) - O_I \end{bmatrix}_{T,I}
\end{aligned} \tag{B.97}$$

A_{r_0,i_0}

for $r_0 = 1 \to R$

$$\begin{aligned}
\mathcal{A}_{0-} &= {}^t\mathbf{W}_{r_0} \left(\tilde{\mathbf{V}}^{\cdot(\beta_0-2)} \otimes \mathbf{V}\right) e^{-\delta_{r_0}\mathbf{M}_{t0}} \\
\mathcal{A}_{0+} &= {}^t\mathbf{W}_{r_0} \tilde{\mathbf{V}}^{\cdot(\beta_0-1)} e^{-\delta_{r_0}\mathbf{M}_{t0}}
\end{aligned}$$

$$A_{r_0,i} \longleftarrow A_{r_0,i} \otimes \frac{\mathcal{A}_{0-}}{\mathcal{A}_{0+}} \tag{B.98}$$

end for

δ_{r_0}

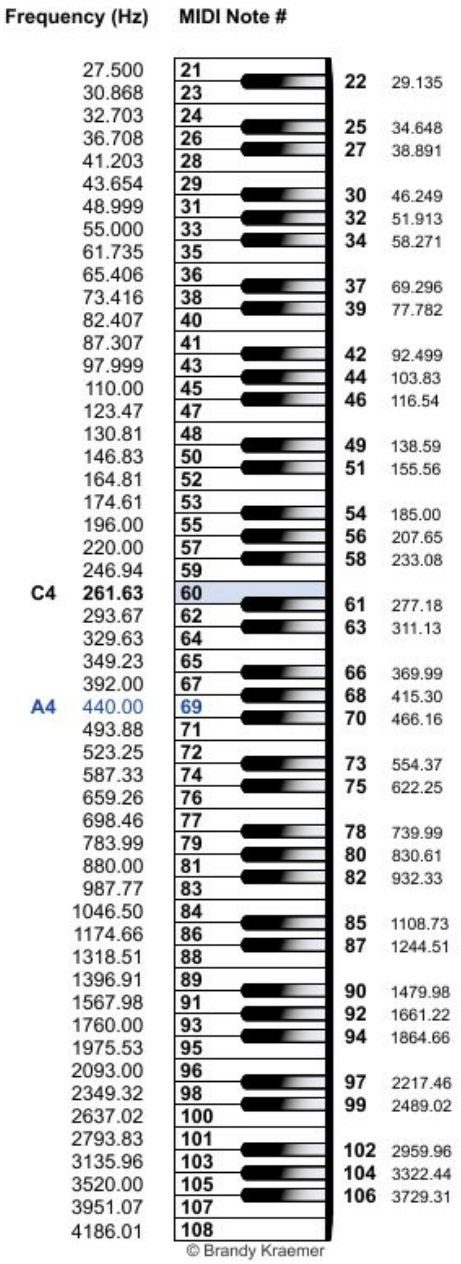
for $r_0 = 1 \to R$

$$\begin{aligned}
\mathcal{D}_{0-} &= {}^t\mathbf{W}_{r_0}\tilde{\mathbf{V}}^{\cdot(\beta_0-1)}\left(\mathbf{M}_{ti}\otimes\mathbf{M}_{to}\otimes e^{-\delta_{r_0}\mathbf{M}_{t0}}\right){}^tA_{r_0,i} \\
\mathcal{D}_{0+} &= {}^t\mathbf{W}_{r_0}\left(\tilde{\mathbf{V}}^{\cdot(\beta_0-2)}\otimes\mathbf{V}\right)\left(\mathbf{M}_{ti}\otimes\mathbf{M}_{to}\otimes e^{-\delta_{r_0}\mathbf{M}_{t0}}\right){}^tA_{r_0,i} \\
\delta_{r_0} &\longleftarrow \delta_{r_0}\otimes\frac{\mathcal{D}_{0-}}{\mathcal{D}_{0+}}
\end{aligned} \tag{B.99}$$

end for

Annexe C

Norme *MIDI Note Number*



Annexe D

Base d'évaluation

D.1 Détail des instruments réels et logiciels

Abréviation	Piano réel / logiciel	Instrument	Conditions d'enregistrement
StbgTGd2	The Grand 2 (Steinberg)	Hybride	Par défaut
AkPnBsdf	Akoustik Piano (Native Instruments)	Boesendorfer 290 Imperial	*Preset* "Gregorian"
AkPnBcht	Akoustik Piano (Native Instruments)	Bechstein D 280	*Preset* "Bechstein Bach"
AkPnCGdD	Akoustik Piano (Native Instruments)	Concert Grand D	*Preset* "Production"
AkPnStgb	Akoustik Piano (Native Instruments)	Steingraber 130 (piano droit)	*Preset* "Modern Play Time"
SptkBGAm	The Black Grand (Sampletekk)	Steinway D	"Ambient" (prise à distance)
SptkBGCl	The Black Grand (Sampletekk)	Steinway D	"Close" (prise de proximité)
ENSTDkAm	Piano réel (Disklavier)	Yamaha Mark III (piano droit)	"Ambient" (prise à distance)
ENSTDkCl	Piano réel (Disklavier)	Yamaha Mark III (piano droit)	"Close" (prise de proximité)

Table D.1 – Liste des instruments réels et logiciels utilisés, et conditions d'enregistrement

D.2 Détail des morceaux

D.2.1 Base de développement

Abréviation	Compositeur	Titre	Instrument
waldstein 3	Beethoven	Sonata No. 21 in C major (Waldstein) , Opus 53 (1804), 3° mvt	ENSTDkCl
alb esp 2	Albéniz	Suite española (1886), Cataluña (Curranda)	SptkBGCl
alb esp 5	Albéniz	Suite española (1886), Asturias (Leyenda)	ENSTDkCl
chpn p2	Chopin	Préludes, Opus 28 (1838), N°2	SptkBGCl
chpn p7	Chopin	Préludes, Opus 28 (1838), N°7	ENSTDkCl
gra esp 3	Granados	Danzas españolas (1900), N°3 - Zarabanda	SptkBGCl
muss 3	Mussorgsky	Pictures at an Exhibition (1874), Promenade - The Tuileries	ENSTDkCl
scn16 4	Schumann	Kreisleriana, Opus 16 (1838), 4° mvt	SptkBGCl
ty februar	Tschaikowsky	The Seasons, Opus 37a (1876), February - Carnival	ENSTDkCl
ty november	Tschaikowsky	The Seasons, Opus 37a (1876), November - Troika Ride	SptkBGCl

TABLE D.2 – Liste des morceaux sélectionnés pour constituer la base de développement.

D.2.2 Base de test

Abréviation	Compositeur	Titre	Instrument
grieg butterfly	Grieg	Lyric Pieces Book III, Opus 43 (1886), N°1 - Butterfly	StbgTGd2
bor ps6	Borodin	Petite Suite (1885), Serenade	ENSTDkCl
chpn op10 e12	Chopin	Etudes, Opus 10 (1832), N°12 - Revolutionary	AkPnCGdD
mendel op62 5	Mendelssohn	Songs without Words Book 5, Opus 62 (1844), N°5 - Venetian Gondola Song	SptkBGAm
ty november	Tschaikowsky	The Seasons, Opus 37a (1876), November - Troika Ride	ENSTDkAm
gra esp 2	Granados	Danzas españolas (1900), N°2 - Oriental	AkPnStbg
deb menu	Debussy	Suite bergamasque (1905), Menuet	StbgTGd2
schub d760 3	Schubert	Fantasia C major (Wanderer) , D 760, Opus 15 (1822), 3° mvt	ENSTDkAm
mz 332 2	Mozart	Sonata No. 12 F major, KV 332 (1783), 2° mvt	StbgTGd2
shumm 1	Schubert	6 Moments musicaux, D 780, Opus 94 (1828), N°1	ENSTDkAm
scn 15 3	Schumann	Scenes from Childhood, Opus 15 (1838), N°3 Blindman's Buff	AkPnBcht
ty januar	Tchaikovsky	The Seasons, Opus 37a (1876), January - At the Fireside	AkPnStgb
liz rhap10	Liszt	19 Hungarian Rhapsodies (1885), N°10	AkPnBsdf
hay 40 1	Haydn	Piano Sonata in G major, Hoboken XVI :40 (1784), 1° Mvt	AkPnBcht
waldstein 1	Beethoven	Sonata No. 21 C major (Waldstein) , Opus 53 (1804), 1° Mvt	AkPnBsdf

TABLE D.3 – Liste des morceaux sélectionnés pour constituer la base de test.