Séparation de signaux audio : principes statistiques de l'analyse en composantes indépendantes et applications au signal monophonique

Emmanuel Vincent

Rapport de stage de DEA ATIAM
Équipe analyse-synthèse, IRCAM
sous la direction de Xavier Rodet

en collaboration avec Bertrand Delezoide

30 juin 2001

Table des matières

Table des figures

Remerciements

Je tiens à remercier particulièrement Xavier Rodet pour m'avoir proposé de travailler sur ce sujet passionnant, ainsi que pour ses remarques et conseils avisés au long de ce stage.

Merci beaucoup à Bertrand Delezoide pour ses recherches parallèles aux miennes et pour les nombreuses discussions que nous avons eues ensemble.

Merci aussi à Jean-François Cardoso et à Shlomo Dubnov pour leur passage à l'IRCAM et nos échanges de points de vue très intéressants.

Merci enfin aux thésards et aux développeurs de l'équipe analyse-synthèse pour leurs discussions et leur bonne humeur, particulièrement lors des pauses.

1 Introduction

J'ai travaillé en collaboration avec Bertrand Delezoide sur le sujet de la séparation de sources sur des signaux monophoniques par analyse en sous-espaces indépendants sur le spectrogramme.

Après avoir étudié à la fois un modèle s'appliquant aux sons stationnaires et la question de l'inversion des représentations temps-fréquence, j'ai écrit les programmes MATLAB correspondants. J'ai effectué alors plusieurs tests sur des mélanges monophoniques, qui m'ont permis de dégager l'importance d'une interprétation non statistique du modèle. Je me suis finalement intéressé à améliorer les résultats du modèle en essayant de le modifier ou de lui ajouter des connaissances a priori.

De son côté, Bertrand a étudié les questions du regroupement des sources suite à l'application du modèle et de l'utilisation du modèle pour des signaux non stationnaires. Il a écrit les programmes MATLAB correspondants, et fait plusieurs tests pour déterminer les mélanges les plus adaptés à cette technique de séparation. Il a aussi développé une approche par filtrage de l'inversion des représentations temps-fréquence des sources. Dans la suite, je signale ses contributions par la référence [Del01].

La **séparation aveugle de sources** (blind source separation, BSS) [BSS] désigne les techniques visant à retrouver des signaux inconnus appelés sources à partir de plusieurs observations de leur mélange, dont les caractéristiques ne sont pas données. Ce problème reste très difficile, surtout lorsque le mélange varie, et en présence de délais, de filtrages et d'échos. On parle alors de **déconvolution aveugle de sources**.

En principe, ces techniques ne s'appliquent pas au signal monophonique, c'est-à-dire dans le cas d'une seule observation (microphone). Cependant, l'**analyse de scènes auditives** [Bre90], dont le but est de décrire des événements sonores divers et concomitants, permet de définir un concept de source dans ce cas et de démixer. Dans ce travail, nous cherchons à retrouver les sources sous-jacentes à des signaux monophoniques par des techniques calculatoires formalisées et relativement simples, en évitant les heuristiques complexes.

Face aux techniques plus classiques d'analyse sinusoïdale ou de corrélogrammes, l'**analyse en composantes indépendantes** (independent component analysis, ICA) [Car98b] est une technique prometteuse pour la BSS, n'exploitant que l'hypothèse d'indépendance entre les sources. Elle a été appliquée avec succès aux sons, aux images [Hyv00], aux signaux médicaux [Lat96], aux communications [Tor98], ou encore aux mathématiques financières. Limitée en théorie à l'analyse d'un nombre d'observations supérieur à celui de sources, l'ICA a été utilisée dans le cas d'une seule observation : l'**analyse en sous-espaces indépendants** (independent subspace analysis, ISA) de spectrogrammes [Cas00] permet maintenant d'étudier les sons monophoniques.

Dans la suite, nous nous intéressons exclusivement à la séparation de sources sur des signaux monophoniques, en exposant un modèle d'ISA sur le spectrogramme. Nous étudions ensuite des améliorations possibles, et nous proposons d'autres modèles.

2 Le modèle

Le modèle que nous allons exposer repose sur l'équivalence entre le signal étudié et une **représentation temps-fréquence** (time-frequency distribution, TFD) de ce signal. Séparer le signal en plusieurs sources se ramène donc à séparer cette TFD en plusieurs TFDs **indépendantes** censées les représenter. Pour donner un sens à la notion d'indépendance dans ce cadre, il est nécessaire d'expliciter plus avant nos hypothèses [Cas98][Cas00][Sma01].

Le signal $x(t)$ monodimensionnel observé est transformé en sa TFD $\mathbf{X}$ en prenant le module d'une **transformée de Fourier à court terme** (TFCT). Bien qu'on ne garde que le module, l'information contenue dans $\mathbf{X}$ permet bien de reconstituer $x(t)$ (cf. 3.6), et les coefficients de la TFD s'interprètent alors comme l'amplitude d'une fréquence donnée à un moment donné. Dans la suite, on suppose que $\mathbf{X}$ est une matrice de taille $m \times n$ (m canaux fréquentiels et n trames temporelles).

Première hypothèse : $\mathbf{X}$ est la superposition d'un nombre inconnu *a priori* de TFDs, censées représenter les sources :

$$\mathbf{X} = \sum_{i=1}^{\rho} \mathbf{X}_i + \sum_{j=1}^{\kappa} \mathbf{R}_j, \tag{1}$$

où les $\mathbf{X}_i$ sont ρ TFDs indépendantes et les $\mathbf{R}_j$ sont κ TFDs de signaux de bruit. Observons que la superposition de TFDs est une opération linéaire dans le plan temps-fréquence. Sous l'hypothèse que la TFD inverse conduit à la superposition correspondante des signaux, l'équation (1) correspond à la représentation dans le plan temps-fréquence de l'équation temporelle de séparation de sources.

Seconde hypothèse : chaque TFD $\mathbf{X}_i$ est composée du produit d'un vecteur de **caractéristique spectrale** $\mathbf{y}_i \in \mathbb{R}^m$ et d'un vecteur de **poids temporel** correspondant $\mathbf{v}_i \in \mathbb{R}^n$:

$$\mathbf{X}_i = \mathbf{y}_i \mathbf{v}_i^T, \tag{2}$$

où les $\mathbf{y}_i$ forment une famille orthonormée de $\mathbb{R}^m$. Cette hypothèse signifie que chaque TFD source est la projection de la TFD observée sur un sous-espace de l'espace des fréquences de la TFD. Notons que $\mathbf{X}_i$ a un spectre constant dans le temps. Ce modèle ne peut donc s'appliquer qu'aux signaux stationnaires. Un modèle semblable existe pour les signaux non stationnaires, en supposant ces hypothèses vérifiées sur de courtes portions de signal [Cas00][Del01].

En représentant les composants de base $\mathbf{y}_i$ et $\mathbf{v}_i$ comme les colonnes de deux matrices $\mathbf{Y}_\rho$ et $\mathbf{V}_\rho$ respectivement, on arrive au modèle matriciel

$$\mathbf{X} = \mathbf{Y}_\rho \mathbf{V}_\rho^T + \mathbf{R}, \tag{3}$$

où $\mathbf{R}$ est la somme des TFDs de bruit, $\mathbf{Y}$ une matrice $m \times \rho$ et $\mathbf{V}$ une matrice $n \times \rho$.

On peut maintenant donner un sens à l'indépendance des TFDs $\mathbf{X}_i$ en la définissant comme l'indépendance des vecteurs de caractéristiques spectrales $\mathbf{y}_i$ correspondants. Pour cela, on fait une dernière hypothèse : chaque $\mathbf{y}_i$ est une suite d'observations de variables aléatoires i.i.d. (indépendantes et identiquement distribuées). On peut alors utiliser le formalisme statistique, au sein duquel la notion d'indépendance est bien définie, en associant à l'ensemble des $\mathbf{y}_i$ un vecteur aléatoire, dont on connaît approximativement la distribution grâce aux coefficients des $\mathbf{y}_i$. L'indépendance des $\mathbf{y}_i$ est alors définie comme l'indépendance des marginaux de ce vecteur aléatoire.

On peut remarquer qu'un traitement statistique est difficilement justifiable dans ce cas, et que le modèle ne tient pas compte de l'ordre des coefficients des $\mathbf{y}_i$. Cependant, il s'agit des hypothèses de l'ICA classique faites sur les $\mathbf{y}_i$ considérés comme les signaux de mélange à démixer. Même lorsque le modèle n'est pas tout à fait valable, l'ICA classique donne de bons résultats, et nous pouvons espérer qu'il en sera de même ici. Nous proposons pour plus de clarté une justification non statistique de l'ICA dans la partie 5.

3 L'ISA sur le spectrogramme

3.1 Étapes algorithmiques

Partant d'un signal temporel $x(t)$, nous calculons (fig. 1) le module de sa TFCT $\mathbf{X}$. Une décomposition en valeurs singulières (cf. 3.3) conduit à l'équation $\mathbf{X} = U_\rho \Sigma_\rho V_\rho^T + R$, où les colonnes de U_ρ sont des caractéristiques spectrales décorrélées. Nous utilisons alors une ICA classique (cf. 3.4) sur ces colonnes, ce qui a pour effet de multiplier U_ρ par une matrice orthogonale Q, de sorte que les colonnes de $U_\rho Q$ soient les plus indépendantes possible. Avec les notations du modèle, on a alors $\mathbf{Y}_\rho = U_\rho Q$, $\mathbf{V}_\rho = V_\rho \Sigma_\rho Q$ et $\mathbf{R} = R$.

3.2 La TFCT

La TFCT (fig. 2) se calcule de la façon suivante :

$$X(rI, k) = \sum_{t=-\infty}^{+\infty} x(t) w(rI - t)\, e^{-\frac{2\pi j k t}{N}}, \tag{4}$$

où w est la fenêtre utilisée pour le lissage de la transformée de Fourier et I le rapport de sous-échantillonage. Lorsqu'on utilise une fenêtre de Hanning, I peut être pris égal au quart de la longueur de la fenêtre.

Implémentation en MATLAB :
Nous utilisons la fonction `specgram` de MATLAB pour calculer le spectrogramme, puis nous prenons le module de ce spectrogramme.

Paramètres :

- fréquence d'échantillonnage $f_e = 44100 Hz$
- fenêtre de Hanning de largeur $nwin = 512$ *éch* $= 12$ *ms*
- nombre de points fréquentiels $nfft = 1024$ *éch*
- overlap entre deux fenêtres successives $noverlap = 384$ *éch* (soit une trame tous les 3 *ms*)

3.3 La SVD

La **décomposition en valeurs singulières** (singular value decomposition, SVD) permet de séparer une matrice en plusieurs composantes orthogonales (autrement dit décorrélées du point de vue statistique). Une matrice réelle S de taille $m \times n$ se décompose ainsi :

$$S = U \Sigma V^T, \tag{5}$$

où Σ est une matrice diagonale de taille $m \times n$ à coefficients réels positifs (généralement classés par ordre décroissant) et U et V sont deux matrices réelles

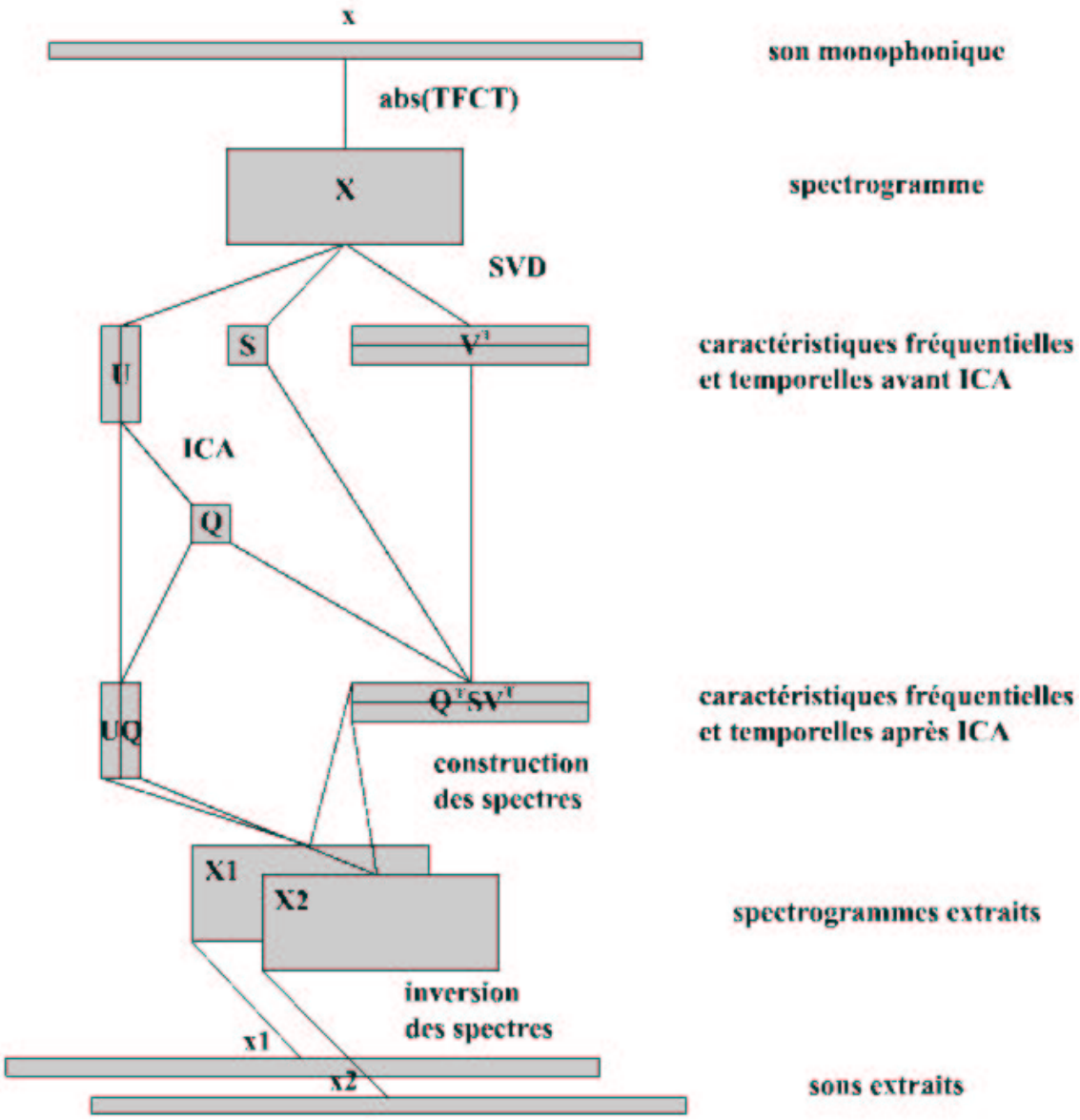

FIG. 1 – Étapes algorithmiques de l'ISA

orthogonales de tailles respectives $m \times m$ et $n \times n$, et où ^{T} dénote la transposition.

Si on décompose une représentation temps-fréquence S, les colonnes u_i de U représentent des caratéristiques fréquentielles de S décorrélées deux à deux, et les colonnes v_i de V des caractéristiques temporelles de S aussi décorrélées deux à deux [Cas98]. L'écriture

$$S = \sum_{i=1}^{min(n,m)} \sigma_{ii} u_i v_i^T \tag{6}$$

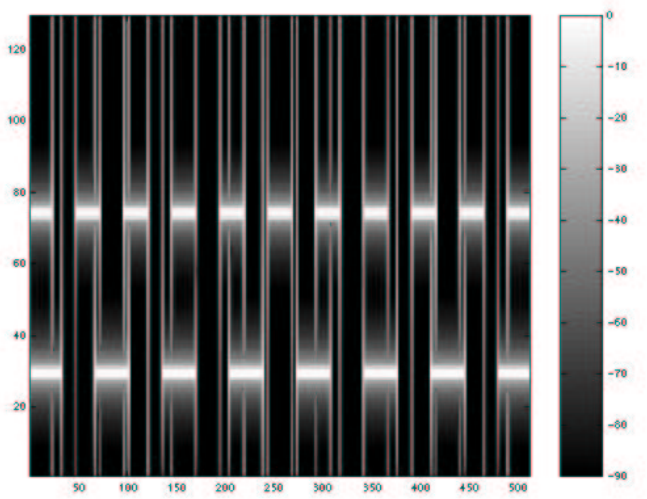

FIG. 2 – Spectrogramme d'un signal d'entrée composé de deux sons purs présents périodiquement

montre que la représentation temps-fréquence s'écrit sous forme de somme de représentations de type $u_i v_i^T$ "décorrélées" au sens où on l'entend ci-dessus, auxquelles on associe une amplitude σ_{ii}. Chaque caractéristique fréquentielle u_i correspond donc à une caractéristique temporelle v_i et à une énergie σ_{ii}^2 (fig. 3, 4, 5).

Pour simplifier les calculs dans la suite, on complète la SVD par la réduction du nombre de composantes mises en jeu. Autrement dit, on ne conserve qu'un certain nombre de caractéristiques fréquentielles et temporelles, comptant pour la plus grande partie de l'énergie du signal, par exemple 99% (fig. 6). Cela revient à considérer maintenant

$$S = U_\rho \Sigma_\rho V_\rho^T, \tag{7}$$

où Σ_ρ est la matrice carrée diagonale contenant les ρ premières colonnes et lignes de Σ, et U_ρ et V_ρ sont les matrices contenant les ρ premières colonnes de U et V, où ρ vérifie

$$\frac{\sum_{i=1}^{\rho} \sigma_{ii}^2}{\sum_{i=1}^{min(n,m)} \sigma_{ii}^2} \geq 99\% \tag{8}$$

Implémentation en MATLAB :
La SVD est réalisée par la fonction MATLAB `svd`.

Paramètres :
Aucun paramètre à régler, si ce n'est qu'on convient de garder une certaine proportion de l'énergie du signal.

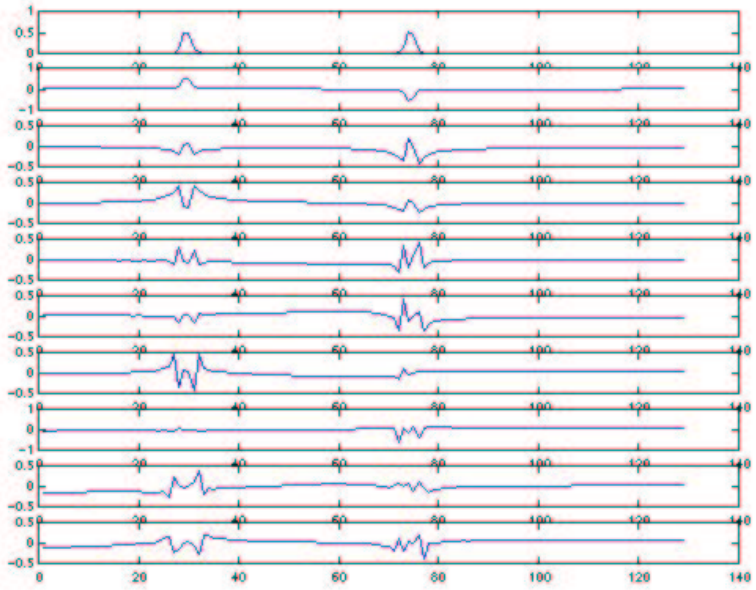

FIG. 3 – Les dix premières caractéristiques spectrales issues de la SVD du signal de la figure 2

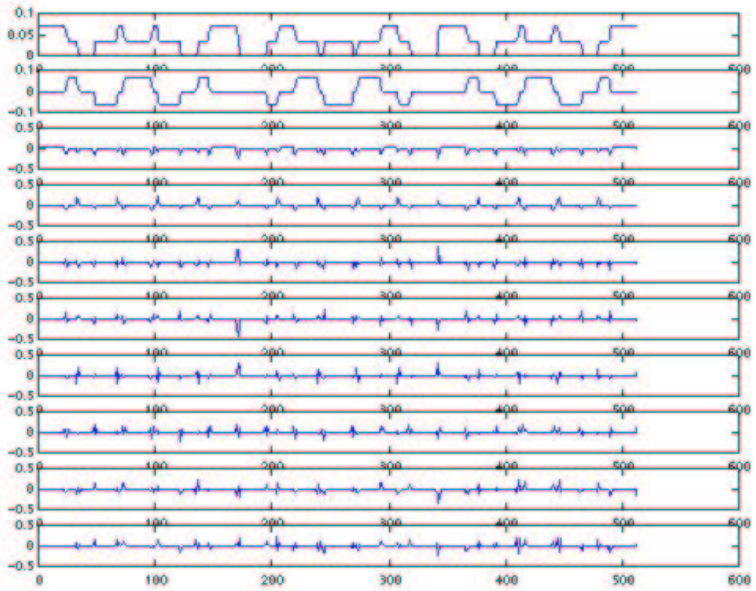

FIG. 4 – Les poids temporels correspondants (signal de la figure 2)

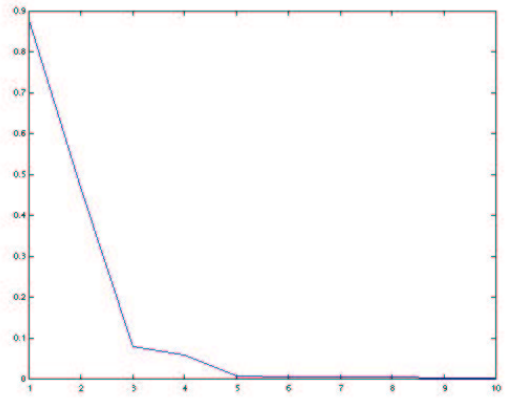

FIG. 5 – Les amplitudes associées (signal de la figure 2)

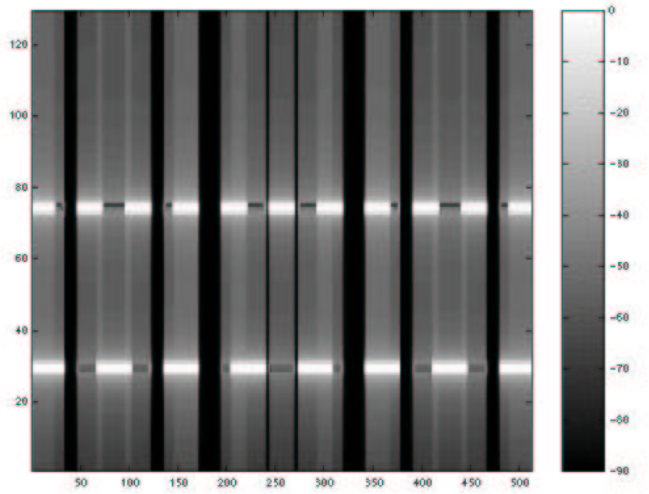

FIG. 6 – Le spectrogramme du signal de la figure 2 après avoir conservé deux composantes (pour 99.0% de l'énergie)

3.4 L'ICA

Considérons n signaux $s_1(t), \ldots, s_n(t)$ indépendants statistiquement appelés sources. Rappelons brièvement que l'on fait l'hypothèse que chaque source est formée d'observations de variables aléatoires i.i.d., et qu'on peut donc utiliser le formalisme statistique et la notion d'indépendance. Le principe de l'ICA est de retrouver ces sources à partir de n observations $x_1(t), \ldots, x_n(t)$ de leur mélange, supposé linéaire, instantané et inversible. Dans le cadre du modèle que nous étudions, les signaux observés sont les colonnes de la matrice U_ρ issue de la SVD (fig. 7).

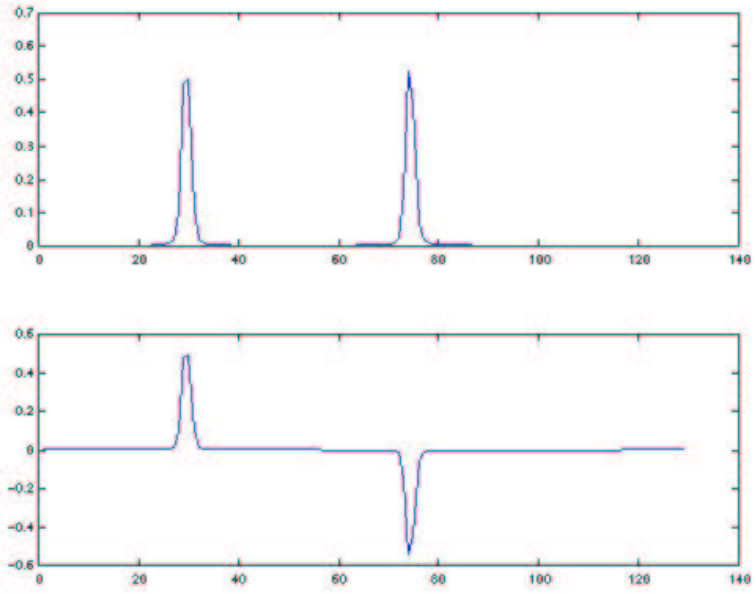

FIG. 7 – Les deux premières caractéristiques spectrales issues de la SVD : les signaux de départ pour l'ICA (signal de la figure 2)

Le modèle s'exprime ainsi sous forme vectorielle [Car98b] :

$$\mathbf{x}(t) = A\mathbf{s}(t). \tag{9}$$

A est appelée matrice de mixage. Le problème se ramène donc à calculer une matrice de démixage B telle que

$$\mathbf{y}(t) = B\mathbf{x}(t) \tag{10}$$

estime correctement les sources.

L'ICA est comparable à l'**analyse en composantes principales** (principal component analysis, PCA), qui considère le même problème à l'exception près que l'indépendance des sources est ramenée à leur simple décorrélation. Si on écrit les observations dans les lignes d'une matrice X, les solutions de la PCA sont les lignes de la matrice $Y = XV = U\Sigma$ où $X = U\Sigma V^T$ est la SVD de X. Comme nous l'avons expliqué ci-dessus, les premières lignes de Y sont décorrélées deux à deux et forment les composantes principales de X au sens de l'énergie. Les hypothèses de l'ICA étant plus contraignantes que celles de la PCA, on va devoir utiliser d'autres caractéristiques des signaux que leurs statistiques d'ordre 2 [Pha92][Bel95][Com94][Car98b][Lee98].

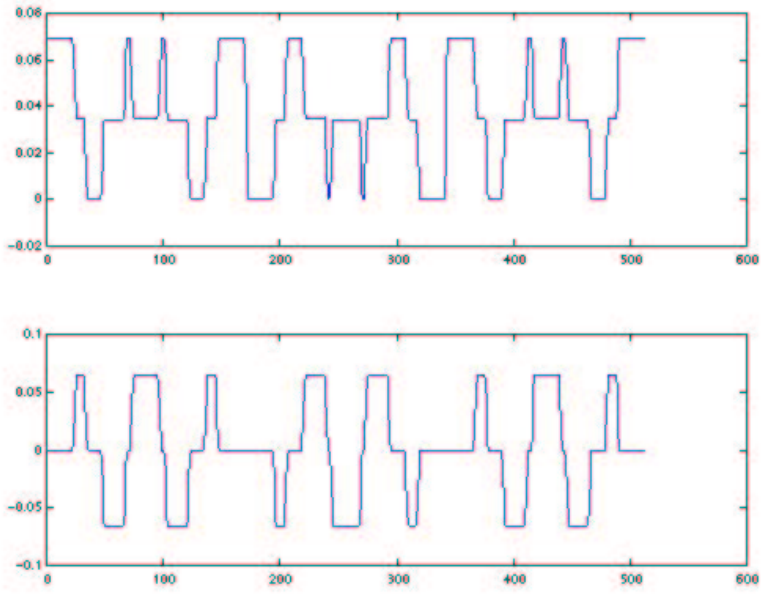

FIG. 8 – Les poids temporels correspondants (signal de la figure 2)

On associe au vecteur $\mathbf{s}(t)$ des sources une densité de probabilité $q(\mathbf{s})$ et au vecteur $\mathbf{x}(t)$ des observations une densité de probabilité $p(\mathbf{x}; A, q)$ dépendant de A et q. Le changement de variable effectué par la matrice A montre que

$$p(\mathbf{x}; A, q) = |detA|^{-1}\, q(A^{-1}\mathbf{x}). \tag{11}$$

Considérons T échantillons des observations $\mathbf{x}(t)$. Ce sont par hypothèse des observations de variables indépendantes. La densité de probabilité d'observation de $\mathbf{x}(t)$ vaut donc $p(\mathbf{x}(1), \ldots, \mathbf{x}(T)) = p(\mathbf{x}(1)) \times \cdots \times p(\mathbf{x}(T))$. On définit la log-vraisemblance normalisée des observations comme

$$\frac{1}{T}\, log\, p(\mathbf{x}(1), \ldots, \mathbf{x}(T); A, q) = \frac{1}{T} \sum_{t=1}^{T} log\, q(A^{-1}\mathbf{x}(t)) - log\, |detA|. \tag{12}$$

À la limite d'un très grand nombre d'échantillons, on obtient :

$$\frac{1}{T}\, log\, p(\mathbf{x}(1), \ldots, \mathbf{x}(T); A, q) \overset{T\to\infty}{\longrightarrow} -\mathbf{K}(A^{-1}\mathbf{x}|\mathbf{s}) + cst, \tag{13}$$

où

$$\mathbf{K}(\mathbf{f}|\mathbf{g}) = \int f(\mathbf{s})\, log \frac{f(\mathbf{s})}{g(\mathbf{s})}\, d\mathbf{s} \tag{14}$$

est la divergence de Kullback entre les variables $\mathbf{f}$ et $\mathbf{g}$ de densités de probabilité f et g.

En particulier, $\mathbf{K}$ est toujours positive et ne s'annule que lorsque les deux densités sont égales, fournissant en quelque sorte une distance non symétrique entre ces densités. La log-vraisemblance des observations est donc à une constante près une mesure de la proximité entre la loi des sources estimées et celle des vraies sources. C'est le principe du **maximum de vraisemblance** (maximum likelihood principle) : lorsque q est fixée, la matrice de mixage A est celle qui maximise la log-vraisemblance des observations , ou de façon équivalente la matrice de démixage B est celle qui minimise la fonction de contraste

$$\phi_{ML}(\mathbf{y}) = \mathbf{K}(\mathbf{y}|\mathbf{s}). \tag{15}$$

Cependant, puisqu'on ne connaît pas les sources, il est impossible d'estimer la fonction de contraste directement. On va donc chercher à la minimiser à la fois par rapport à B et par rapport à $q(\mathbf{s})$. On définit $\tilde{\mathbf{y}}$ comme le vecteur aléatoire dont les marginaux sont indépendants et de même distribution que ceux de $\mathbf{y}$. Sous l'hypothèse d'indépendance des sources, une égalité classique donne :

$$\phi_{ML}(\mathbf{y}) = \mathbf{K}(\mathbf{y}|\tilde{\mathbf{y}}) + \mathbf{K}(\tilde{\mathbf{y}}|\mathbf{s}). \tag{16}$$

Comme $\mathbf{K}(\mathbf{y}|\tilde{\mathbf{y}})$ ne dépend pas de $\mathbf{s}$, on aboutit à une nouvelle fonction de contraste, dite d'**information mutuelle** (mutual information) :

$$\phi_{MI}(\mathbf{y}) = \mathbf{K}(\mathbf{y}|\tilde{\mathbf{y}}). \tag{17}$$

$\phi_{MI}(\mathbf{y})$ est minimale dès que les distributions de $\mathbf{y}$ et $\tilde{\mathbf{y}}$ sont identiques, ce qui veut bien dire que les marginaux de $\mathbf{y}$ sont indépendants.

La minimisation est effectuée par une descente de gradient. Au lieu du gradient classique de la fonction de contraste, on utilise le gradient relatif $\nabla\phi$, défini de la manière suivante :

$$\phi(\mathbf{y} + \varepsilon\mathbf{y}) = \phi(\mathbf{y}) + \langle \nabla\phi(\mathbf{y}) \,|\, \varepsilon \rangle + o(\|\varepsilon\|), \tag{18}$$

où $\langle .|. \rangle$ est le produit scalaire euclidien entre matrices. Ce gradient relatif assure une descente effective de la fonction de contraste à chaque itération, puisque $\phi(\mathbf{y} - \mu\nabla\phi(\mathbf{y})\mathbf{y}) \simeq -\mu\|\nabla\phi(\mathbf{y})\|^2$. De plus, il donne un algorithme stable et avec des performances comparables quel que soit le mélange à inverser [Car98b]. Dans le cas de ϕ_{MI}, en utilisant le fait que les marginaux de $\tilde{\mathbf{y}}$ sont indépendants, le gradient relatif vaut :

$$\nabla\phi_{MI}(\mathbf{y}) = \mathbb{E}\, H_\varphi(\mathbf{y}) \tag{19}$$

où $H_\varphi : \mathbb{R}^n \mapsto \mathbb{R}^{n\times n}$ vaut

$$H_\varphi(\mathbf{y}) = \varphi(\mathbf{y})\mathbf{y}^T - I \tag{20}$$

avec $\varphi : \mathbb{R}^n \mapsto \mathbb{R}^n$ le vecteur colonne contenant toutes les fonctions score des marginaux de $\mathbf{y}$:

$$\varphi_i = -(log\, f_i)' = -\frac{f_i'}{f_i} \tag{21}$$

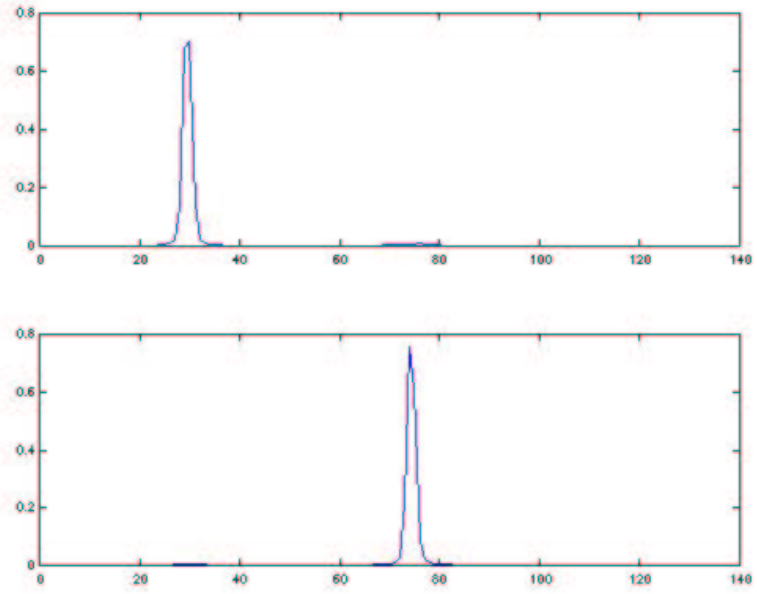

FIG. 9 – Les caractéristiques spectrales à la sortie de l'ICA (signal de la figure 2)

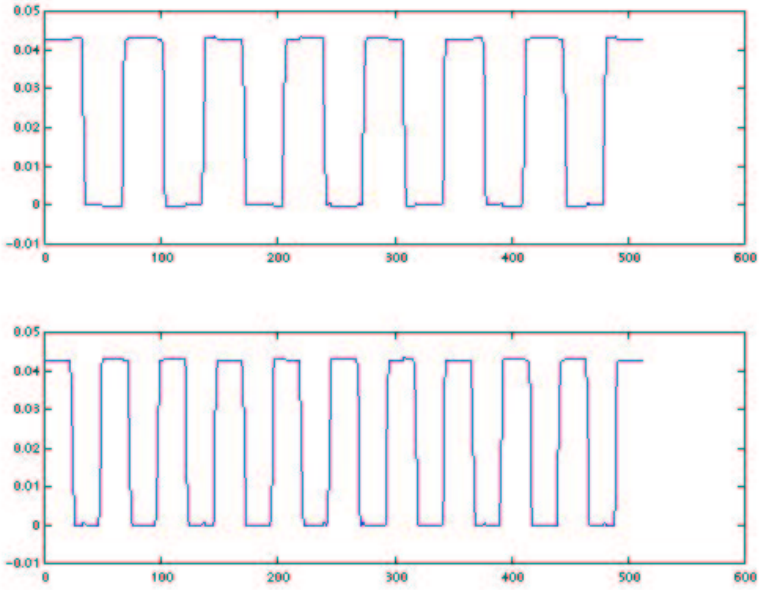

FIG. 10 – Les poids temporels correspondants (signal de la figure 2)

où f_i est la densité de probabilité du marginal i de $\mathbf{y}$.

Dans la suite, on supposera que les observations sont des vecteurs orthonormés (car issus de la SVD). Comme c'est le cas aussi pour les sources (puisqu'on cherche une base orthonormée de l'espace des fréquences de la TFD), on cherche une matrice de démixage orthogonale. On modifie légèrement le gradient pour tenir compte de cette nouvelle contrainte en considérant un nouveau H_φ :

$$H^o_\varphi(\mathbf{y}) = \mathbf{y}\mathbf{y}^T - I + \varphi(\mathbf{y})\mathbf{y}^T - \mathbf{y}\varphi(\mathbf{y})^T. \tag{22}$$

On peut maintenant appliquer un algorithme de gradient de la façon suivante. On prend comme point de départ une matrice B orthogonale quelconque. À chaque étape, on calcule le vecteur $\mathbf{y} = B\mathbf{x}$ des sources estimées, puis on estime $\nabla\phi_{MI}(\mathbf{y})$ en remplaçant l'espérance $\mathbb{E}\, H^o_\varphi(\mathbf{y})$ par une moyenne sur les données $\frac{1}{T}\sum_{t=1}^{T} H^o_\varphi(\mathbf{y}(t))$. On modifie alors B en $(I - \mu\nabla\phi(\mathbf{y}))B$ où μ est le pas de l'algorithme, et on passe à l'itération suivante. Le procédé converge alors vers un minimum local de la fonction de contraste.

Pour appliquer ce procédé il ne reste plus qu'à savoir estimer le vecteur $\varphi(\mathbf{y(t)})$ des fonctions score des sources estimées. Pour cela, on se sert de la définition (21) en estimant la densité de probabilité q_i de y_i par la méthode des noyaux [Ant01]. On part de l'estimateur classique par la méthode de l'histogramme, qui consiste à séparer l'espace des observations en plusieurs tranches et à compter le nombre d'observations dans chaque tranche, et on le lisse par un noyau K approprié :

$$\widehat{q}_i(y) = \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{y - y_i(t)}{h}\right). \tag{23}$$

Le paramètre h sert à régler la largeur du noyau : plus il est élevé, plus la densité estimée est lisse. Lorsque le noyau K satisfait certaines conditions (intégrale unité, moyenne nulle et énergie finie), l'estimateur est ainsi bien meilleur que le simple histogramme. Dans la pratique, on choisit un noyau gaussien, et on se limite à calculer les $\widehat{q}_i(y_i(t))$, qui suffisent à trouver $\varphi(\mathbf{y(t)})$.

Implémentation en MATLAB :
Plusieurs packages MATLAB pour l'ICA comme FastICA [FastICA][Hyv97] ou JADE [JADE][Car99] utilisent une méthode différente d'estimation des fonctions score par des statistiques d'ordre élevé comme les cumulants d'ordre 4, dont nous donnons un aperçu dans la partie 5.

Nous utilisons selon les principes exposés ci-dessus le MIalgorithm [MIalgorithm] [Tal97] de Christian Jutten, que j'ai modifié en `icanopar` pour ne garder que le cœur du programme, utiliser le gradient H^o_φ au lieu de H_φ et partir d'une matrice de démixage quelconque au lieu de l'identité. Ce programme utilise le sous-programme `estim_psi` qui calcule les $\widehat{\varphi}_i(y_i(t))$.

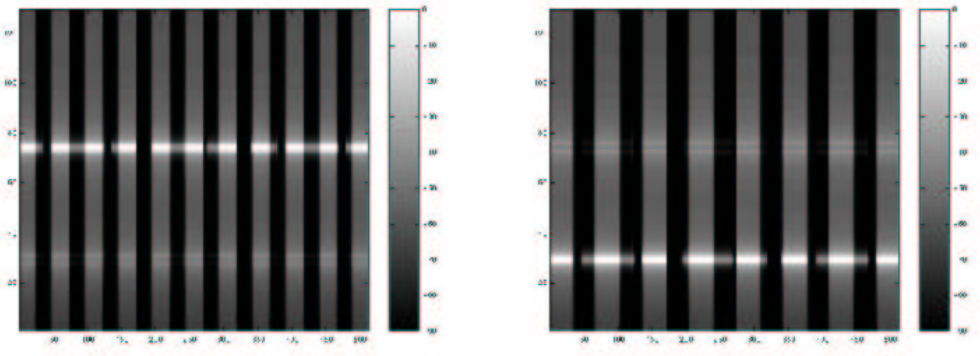

FIG. 11 – Les deux spectrogrammes "indépendants" (signal de la figure 2)

Paramètres :

- pas de descente $\mu = 1$
- noyaux gaussiens de largeur $h = 0.1$ (dans la pratique, la largeur exacte du noyau importe peu mais elle ne doit pas être trop faible : nous avons eu de bons résulats pour des largeurs de noyaux variant de .05 à .2, alors même que les variables considérées sont toutes à valeurs entre -1 et 1)
- critère d'arrêt $errormax = 0.02/T$ (pour T samples) sur le plus grand coefficient de la matrice de gradient, cette erreur étant considérée généralement comme non significative statistiquement
- on observe généralement une convergence rapide de l'algorithme au bout de 20 à 50 itérations, ensuite l'erreur ne diminue quasiment plus

3.5 Construction des TFDs indépendantes

L'application d'une ICA aux colonnes de U_ρ multiplie donc U_ρ par une matrice orthogonale Q, de sorte que les colonnes de $U_\rho Q$ soient les plus indépendantes possible. Les hypothèses du modèle montrent alors que les composantes indépendantes recherchées (fig. 11) sont les matrices de TFD $\mathbf{X}_i = \mathbf{y}_i \mathbf{v}_i^T$, où les $\mathbf{y}_i$ sont les colonnes de $U_\rho Q$ (fig. 9) et les $\mathbf{v}_i$ celles de $V_\rho \Sigma_\rho Q$ (fig. 10).

Cependant, l'ICA ne donne pas toujours les mêmes résultats, dans la mesure où la fonction de contraste ne varie pas lorsqu'on change le signe d'une des caractéristiques spectrales, où lorsqu'on les permute. On rajoute donc des conditions d'unicité de l'ICA avant la construction des TFDs indépendantes. Pour supprimer la possibilité de permutation des composantes, on décide de les classer par énergie décroissante, sachant que l'énergie d'une composante $\mathbf{y}_i \mathbf{v}_i^T$ se mesure par la norme euclidienne de $\mathbf{v}_i$ (dans la mesure où $\mathbf{y}_i$ est de norme 1). Pour rendre unique le signe des composantes, on peut forcer le premier coefficient de chaque $\mathbf{y}_i$ à être positif, ou forcer leur moyenne à être positive, ou toute autre condition adéquate.

3.6 Inversion des TFDs

On veut maintenant inverser les TFDs pour obtenir les signaux temporels résultats. Cependant, on ne peut pas appliquer l'algorithme classique d'overlap-add : on dispose uniquement du module de ces TFDs, et ce ne sont pas forcément des spectrogrammes (c'est-à-dire qu'ils ne sont pas forcémént les modules de TFCTs de signaux temporels).

Basé sur la redondance des représentations temps-fréquence, l'algorithme de Griffin et Lim [Gri84] résout ces deux problèmes : il trouve le signal temporel dont la TFCT est la plus proche au sens des moindres carrés de la représentation à inverser, et il retrouve l'information de phase perdue. Il fonctionne par des projections successives sur l'espace temporel et l'espace fréquentiel.

Soit $Y_0(rI,k)$ la représentation à inverser. On construit une suite de signaux $y_i(t)$ convergeant vers la solution de la façon suivante :

$$y_i(t) = \frac{\sum_{r=-\infty}^{+\infty}\left(\frac{1}{N}\sum_{k=0}^{N-1} Y_i(rI,k)\, e^{\frac{2\pi jkt}{N}}\right) w(rI-t)}{\sum_{r=-\infty}^{+\infty} w^2(rI-t)}, \tag{24}$$

$$Z_{i+1}(rI,k) = \sum_{m=-\infty}^{+\infty} s(m)w(rI-m)\, e^{-\frac{2\pi jkm}{N}}, \tag{25}$$

$$Y_{i+1}(rI,k) = \frac{Z_{i+1}(rI,k)}{|Z_{i+1}(rI,k)|} Y_0(rI,k). \tag{26}$$

Le signal temporel obtenu avec (24) est retransformé en TFCT par (25), et on se sert de l'estimation de phase de cette TFCT et du module de Y_0 pour se ramener à l'inversion d'une nouvelle représentation temps-fréquence par (26). On peut montrer que l'erreur définie par

$$\frac{\sum_{r,k}(|Z_i(rI,k)| - Y_0(rI,k))^2}{\sum_{r,k} Y_0(rI,k)^2} \tag{27}$$

décroît à chaque itération.

Cependant, cet algorithme converge très lentement, et il nécessaire d'estimer une phase initiale non nulle pour obtenir une convergence plus rapide. L'estimation la plus simple [Sla94] consiste à appliquer un retard de phase linéaire optimal à chaque trame de la TFCT. Concrètement, lors de la première itération de l'algorithme, la transformée de Fourier inverse

$$\frac{1}{N}\sum_{k=0}^{N-1} Y_i(rI,k)\, e^{\frac{2\pi jkt}{N}} \tag{28}$$

de chaque trame r de la TFCT subit une rotation en

$$\frac{1}{N}\sum_{k=0}^{N-1} Y_i(rI,k)\, e^{\frac{2\pi jk(t+t')}{N}}, \tag{29}$$

et on choisit le paramètre t' optimal, $0 \leq t' \leq nfft - 1$, conduisant à une corrélation maximale entre la contribution (29) de cette trame r (non encore fenêtrée par w) et la partie du signal déjà reconstituée par l'overlap-add fenêtré des trames r', $r' \leq r$.

Grâce à cette estimation initiale, l'algorithme converge beaucoup plus rapidement (en moyenne 25 itérations au lieu de plusieurs centaines). Notons enfin qu'il est possible dans une optique de test de l'ISA d'utiliser au départ des signaux connus mixés, et de se servir des phases des TFCT de ces signaux comme estimation initiale de phase pour assurer une convergence encore plus rapide et s'assurer que les résultats dépendent des qualités de séparation de l'algorithme d'ISA plutôt que de celui d'inversion des représentations temps-fréquence obtenues. Généralement, nous utilisons aussi la phase du mélange initial comme estimation de départ, avant d'appliquer l'algorithme décrit ci-dessus (qui est indispensable car la phase du mélange diffère souvent un peu de celles des composantes).

Implémentation en MATLAB :
J'ai écrit deux programmes pour cette tâche : le programme `slaney` estime la phase initiale et effectue 25 itérations de l'algorithme de Griffin et Lim, il utilise le programme `grifflim`, qui calcule l'étape de l'algorithme présentée dans l'équation (24).

Paramètres :
On reprend les paramètres de la TFCT.

4 Exemples

4.1 Analyse d'un bruit

Le modèle d'ISA sur le spectrogramme est très utile pour extraire des caractéristiques d'un son produit par un objet unique. On étudie ici le bruit d'un bris de verre (fig. 12). On garde six composantes pour 85.0% de l'énergie (échantillonage à 44100Hz, fenêtre de largeur 256, 256 points de FFT, noyau gaussien de largeur 0.15).

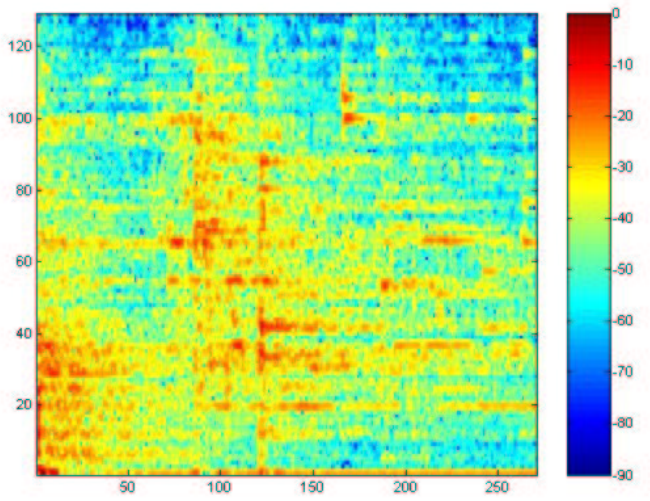

FIG. 12 – Spectrogramme d'un son de bris de verre

On remarque sur le spectrogramme trois caractéristiques intéressantes : un bruit à très basse fréquence lors de l'impact avec une enveloppe décroissant lentement, un bruit à bande assez large lors de l'impact, et plusieurs résonances décroissantes à haute fréquence ensuite, avec chacune une fréquence et un point d'attaque différents. On voudrait représenter ces caractéristiques comme des vecteurs indépendants formant une base spectrale, associés à des poids temporels différents.

Les résultats sont présentés dans la figure 15. On voit que la composante 2 contient l'impact à basse fréquence de l'attaque, et que les composantes 4, 5 et 6 représentent les vibrations à haute fréquence des particules de verre après l'impact. On peut même observer sur la figure 14 la forme des attaques et des décroissances de ces vibrations (en exponentielle décroissante, sans tenir compte du bruit important).
Par contre, le bruit à bande assez large à l'impact n'a pas été détecté convenablement : il est présent dans la composante 1, mais cette composante contient aussi un bruit de spectre identique peu de temps après, et qui était inexistant

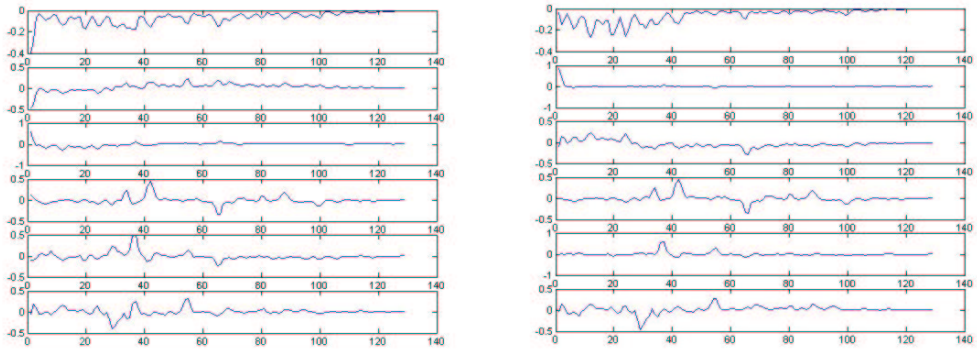

FIG. 13 – Les caractéristiques spectrales avant et après ICA (bris de verre)

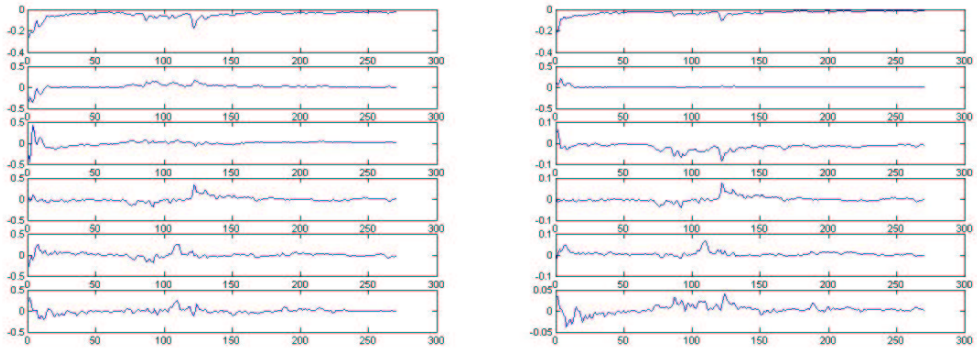

FIG. 14 – Les poids temporels correspondants avant et après ICA (bris de verre)

sur le spectrogramme originel. De fait, ce bruit apparu dans la composante 1 sans exister auparavant s'annule avec le même bruit présent dans la composante 3. En effet, l'observation des caractéristiques fréquentielles et temporelles de ces deux composantes (fig. 13, 14) montre que les coefficients de spectrogramme correspondant à ce bruit apparu sont de signe opposé et s'annulent donc à la reconstruction (cf. 6.1).

On note enfin le rôle essentiel de l'ICA dans l'extraction de ces composantes. À la suite de la SVD, les composantes sont assez bruitées et leur variation temporelle présente des à-coups. Au contraire, les composantes issues de l'ICA ont un spectre et des variations temporelles beaucoup plus lisses (cf. partie 5), correspondant plus à la réalité physique.

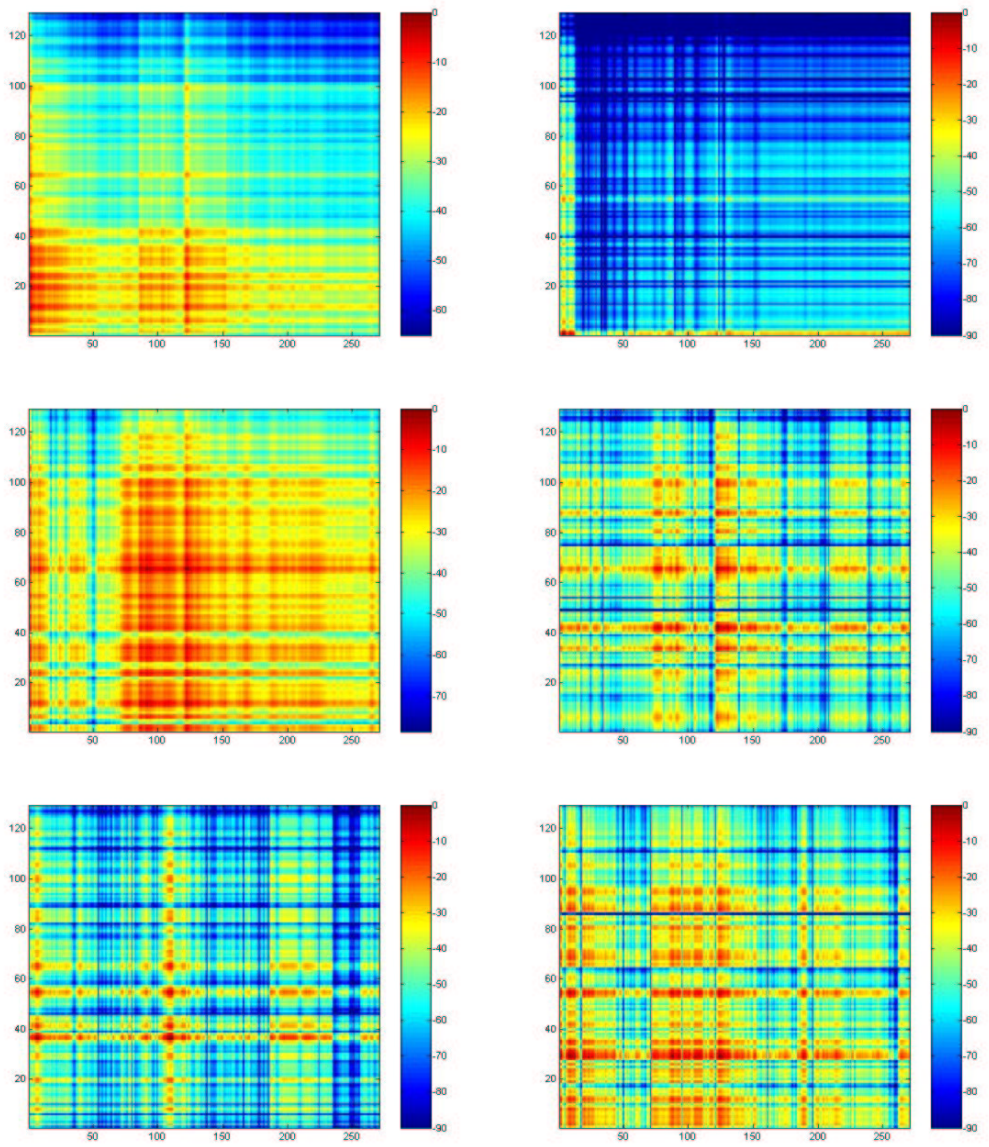

FIG. 15 – Les six TFDs indépendantes extraites du son de bris de verre (classées par énergie décroissante de haut en bas et de gauche à droite)

4.2 Extraction d'attaques

On a constaté avec l'exemple précédent que l'ISA avait extrait l'attaque du bris de verre grâce à son spectre différent de celui du reste du signal. On étudie donc dans cet exemple l'extraction des deux instruments dans un extrait de musique pour tabla et cithare.
Sur le spectrogramme (fig. 16), on distingue bien les caractéristiques de chaque instrument. Les tabla ont un spectre très large et jouent des notes très courtes, avec un faible volume, sauf à deux instants où le volume est plus fort et le

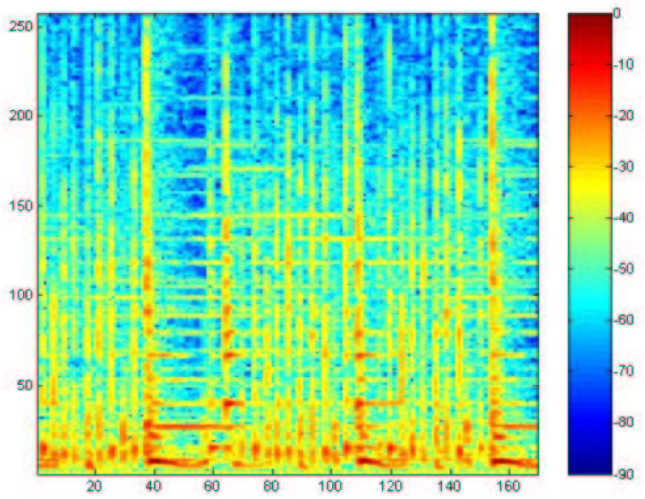

FIG. 16 – Spectrogramme d'un son de tabla et de cithare

spectre un peu différent. La cithare a un spectre assez large à l'attaque, d'où se détachent des harmoniques à la résonance, ces dernières recouvrant en partie des notes de tabla.

On prend les paramètres suivants : échantillonage à 11025Hz, fenêtre de largeur 512, 512 points de FFT, noyau gaussien de largeur 0.1. On prend beaucoup de points de FFT pour avoir une bonne précision de calcul sur la partie utile du signal, la majorité de l'énergie du signal se situant dans les basses fréquences. On garde quatre composantes pour 89.6% de l'énergie.

Les résultats sont présentés dans la figure 19. La composante 1 représente les attaques de la cithare, la composante 2 les notes de faible volume du tabla, la composante 3 les résonances de la cithare, et la composante 4 les deux notes de volume plus fort des tabla.
Encore une fois, on note que l'ICA a nettement lissé les spectres et les variations temporelles des composantes. En particulier, les spectres deviennent tous presque nuls en hautes fréquences, et les notes à rythme rapide des tabla n'apparaissent plus que dans une composante au lieu de deux (fig. 17, 18).

Cependant, on n'a pas a proprement parler séparé les deux instruments, puisque le son de cithare est formé du regroupement des composantes 1 et 3, alors que celui des tabla provient des deux autres composantes. En effet, l'ICA a estimé les composantes les plus indépendantes possible, mais il arrive souvent qu'elles ne soient pas tout à fait indépendantes, ce qui explique que des informations semblables puissent se retrouver dans plusieurs composantes [Car98a]. On peut bien sûr écouter les composantes à la fin de l'algorithme et décider de les regrouper manuellement. Un choix plus judicieux est de les regrouper auto-

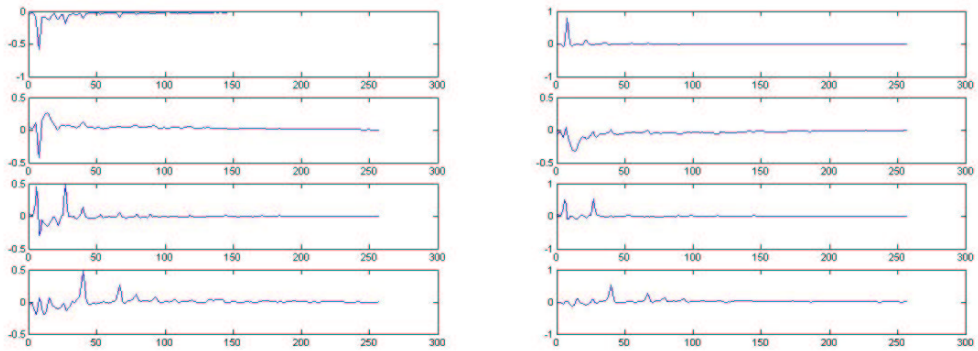

FIG. 17 – Les caractéristiques spectrales avant et après ICA (son de tabla et de cithare)

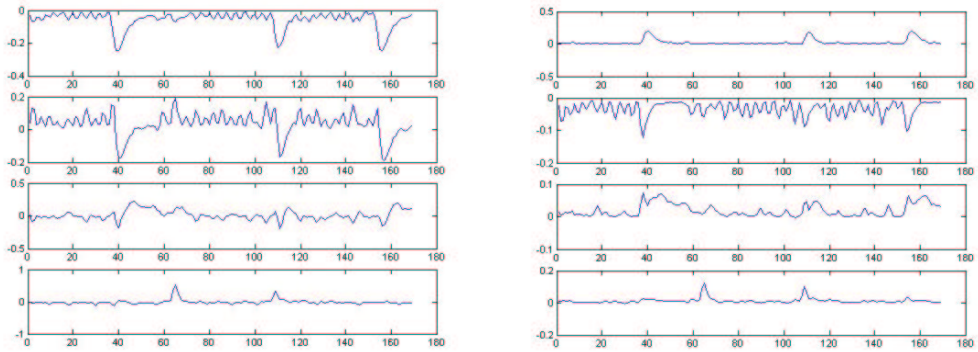

FIG. 18 – Les poids temporels correspondants avant et après ICA (son de tabla et de cithare)

matiquement par un algorithme de regroupement basé sur des distances entre les composantes, faisant appel à leurs caractéristiques fréquentielles [Cas00], ou temporelles, ou les deux [Del01].

Enfin, le principal problème de cet exemple est que les composantes sont extraites avec beaucoup d'imprécision. Par exemple, le volume des tabla est renforcé aux instants des attaques de la cithare, et les résonances de la cithare sont aussi détectées sur les notes de volume élevé des tabla. Utiliser plus de composantes améliore un peu le résultat, mais augmente le temps de calcul, d'autant plus qu'un regroupement devient indispensable à la fin. Dans tous les cas, les signaux audio reconstitués à la fin de l'algorithme ne restituent pas vraiment le son des deux instruments. Un algorithme de reconstitution par filtrage variant dans le temps du son initial par des filtres dépendant des composantes trouvées [Del01] peut donner de très bons résultats sans augmenter

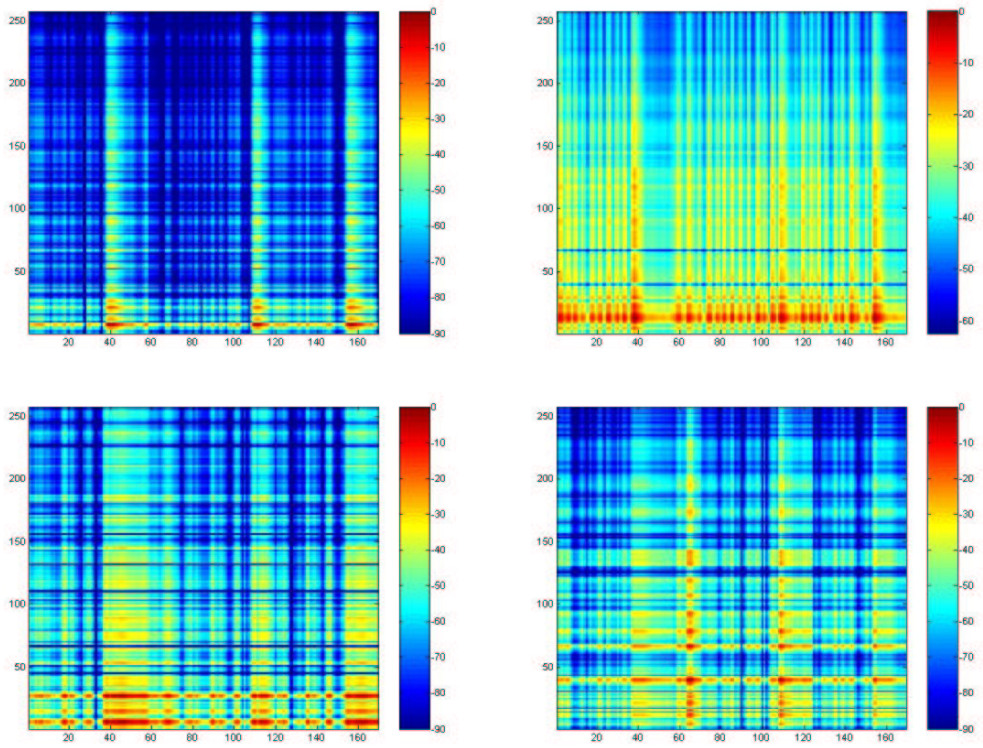

FIG. 19 – Les quatre TFDs indépendantes extraites (son de tabla et de cithare)

le temps de calcul.

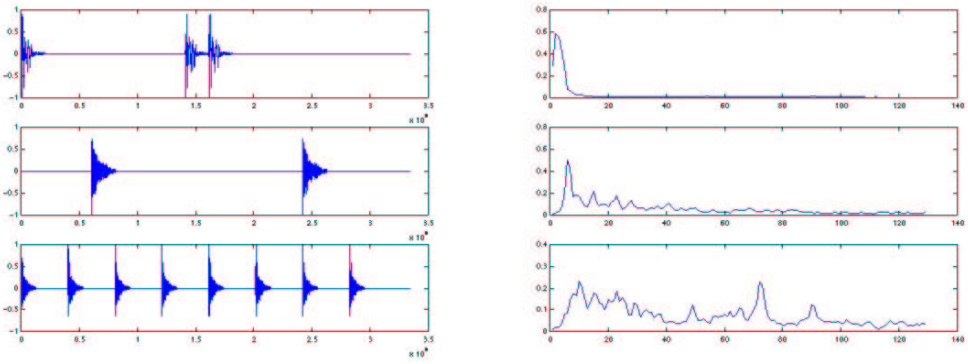

FIG. 20 – La "bass drum", la "snare drum" et le "hi-hat", et les composantes principales de leurs spectres

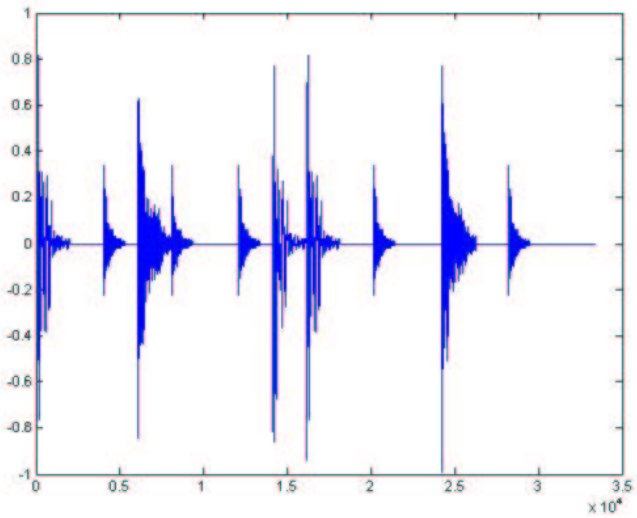

FIG. 21 – Le signal constitué du mélange des trois instruments

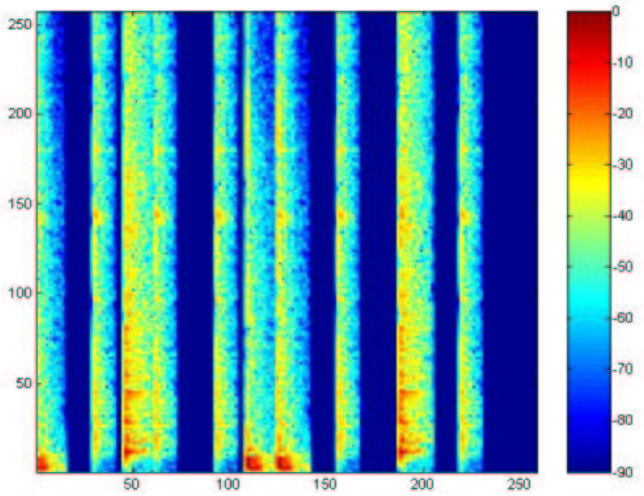

FIG. 22 – Spectrogramme de l'extrait de percussions

4.3 Séparation de sources sur un extrait de percussions

Nous finissons cette série d'exemples par la séparation d'instruments d'un extrait de percussions (fig. 21, 22) contenant trois instruments : une "bass drum", une "snare drum" et un "hi-hat". On peut voir sur la figure 20 leurs formes d'onde (que j'ai extraites à la main à partir des occurences isolées de chaque instrument) et leurs spectres (ou plutôt les composantes principales des SVDs de leurs spectrogrammes en module). La "bass drum" a un spectre bien

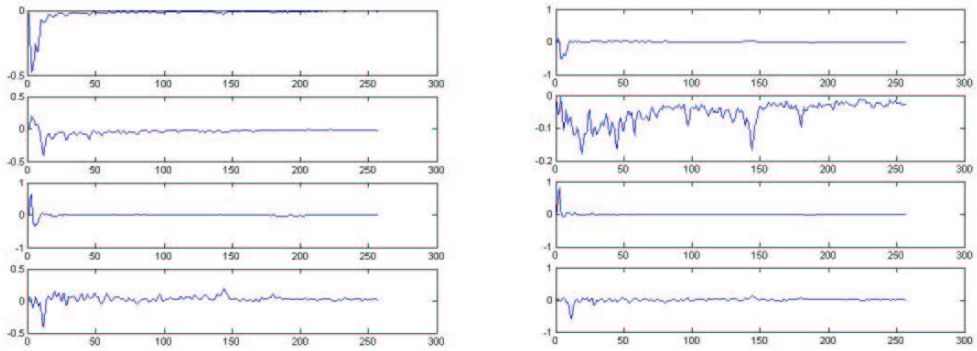

FIG. 23 – Les caractéristiques spectrales avant et après ICA (extrait de percussions)

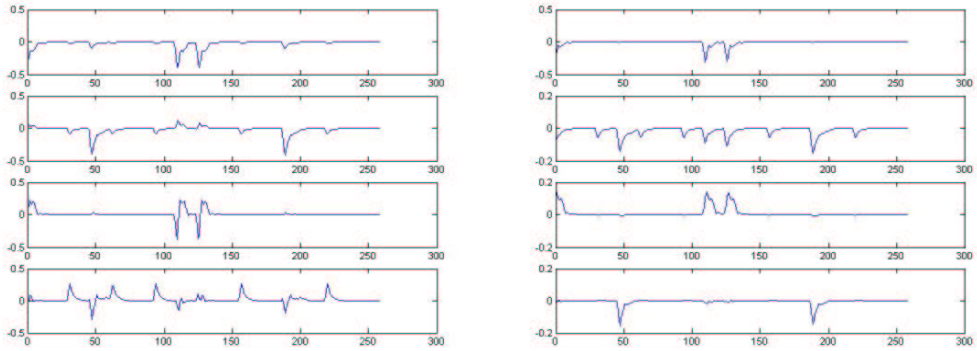

FIG. 24 – Les poids temporels correspondants avant et après ICA (extrait de percussions)

séparé des autres, et sa deuxième occurence est isolée. La "snare drum" a une fréquence de résonance et un spectre à bande assez large, seule sa première occurence est isolée. Enfin, le "hi-hat" a un spectre à bande très large avec des hautes fréquences résonantes. Ses occurences sont souvent cachées par des occurences des deux autres instruments, et son spectre est partiellement recouvert par celui de la "snare drum".

On prend les paramètres suivants : échantillonage à 11025Hz, fenêtre de largeur 512, 512 points de FFT (toujours beaucoup d'énergie en basses fréquences), noyau gaussien de largeur 0.1. On garde quatre composantes pour 94.7% de l'énergie.

Les résultats sont présentés dans les figures 23, 24 et 25. Sans surprise, la "bass drum" a été correctement extraite grâce à son spectre disjoint de celui

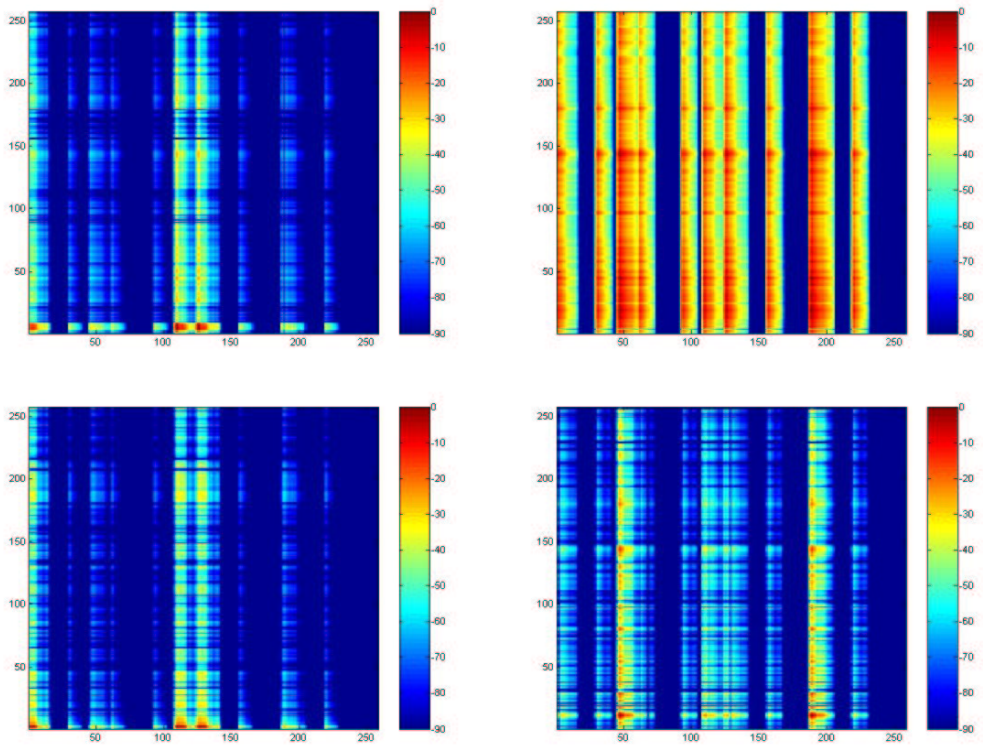

FIG. 25 – Les quatre TFDs indépendantes séparées de l'extrait de percussions

des deux autres instruments (composantes 1 et 3). La "snare drum" apparaît dans la composante 4, un peu moins bien extraite à cause d'une légère ressemblance avec le spectre du "hi-hat". Enfin, le "hi-hat" n'est pas du tout extrait. Le spectre de la composante 2 lui ressemble en hautes fréquences, mais il contient aussi beaucoup de basses fréquences, ce qui explique que cette composante contienne à la fois des occurences du "hi-hat" et des occurrences des deux autres instruments.
Cela est dû a plusieurs raisons : son spectre partiellement recouvert par celui de la "snare drum", mais aussi son faible volume sonore et la prépondérance des basses fréquences dans le signal (cf. la composante principale du spectre trouvée par la SVD dans la figure 23). Les résultats restent semblables en augmentant le nombre de composantes. Par exemple, avec six composantes, on obtient trois composantes pour la "bass drum", deux pour la "snare drum", et une pour le reste.

Le faible volume sonore du "hi-hat" est visiblement une des causes principales de sa mauvaise séparation. On voit sur les figures 26 et 27 les résultats de la SVD et de l'ICA sur le même son lorsqu'on multiplie l'amplitude du "hi-hat" par 1.5. Le spectre de la composante 1 ressemble beaucoup plus à celui du "hi-hat", en particulier il contient peu de basses fréquences comparé à celui de la

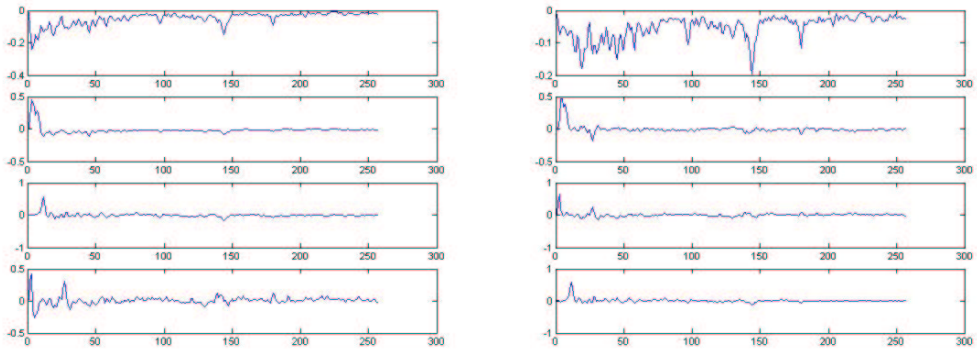

FIG. 26 – Les caractéristiques spectrales avant et après ICA, en multipliant l'amplitude du "hi-hat" par 1.5 dans l'extrait de percussions

composante 4 de la figure 23. En regardant le poids temporel de cette composante, on voit qu'elle confond moins les occurrences du "hi-hat" avec celles des deux autres instruments. Cependant, la "snare drum" est un peu moins bien extraite, toujours à cause de son léger recouvrement spectral avec le "hi-hat". On obtient des résultats similaires pour des amplitudes du "hi-hat" variant de 1.2 à 1.8 environ. En-dessous, il n'est pas bien séparé, et au-dessus les autres instruments ne sont pas bien séparés.

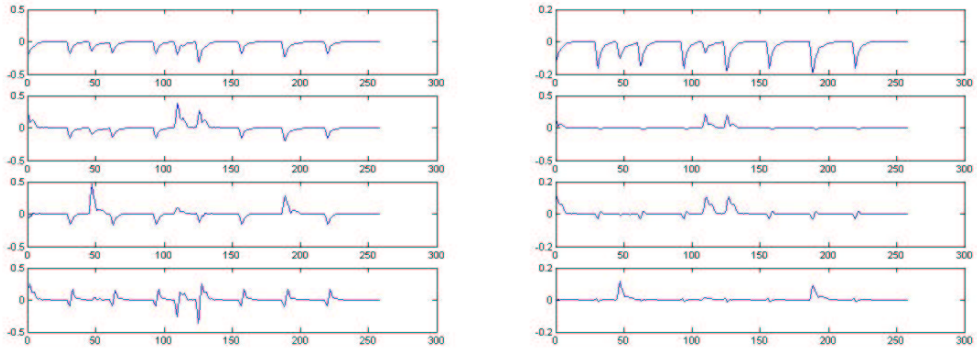

FIG. 27 – Les poids temporels correspondants avant et après ICA (extrait de percussions avec amplitude du "hi-hat" multipliée par 1.5)

Dans la suite, on cherchera à améliorer la séparation des instruments de cet extrait en modifiant le modèle d'ISA sur le spectrogramme, sans modifier le volume sonore des composantes ni leur nombre.

5 Interprétation non statistique de l'ICA

L'ISA sur le spectrogramme a le défaut d'utiliser un modèle probabiliste assez éloigné de la réalité. Pour mieux justifier le recours à l'ICA, on en propose ici une interprétation non probabiliste : l'ICA tend a diminuer au maximum la largeur du support des composantes indépendantes estimées [Car01].

Dans la suite, on utilise les notations de la partie 3.4. Commençons par définir quelques bases statistiques nécessaires et poser quelques notations. Étant donné un vecteur de variables aléatoires centrées $\mathbf{y}$, on définit ses cumulants d'ordre 2 et 4 comme :

$$\left\{ \begin{array}{lcl} \mathcal{C}_{ij}(\mathbf{y}) & = & \mathbb{E}\, y_i y_j \\ \mathcal{C}_{ijkl}(\mathbf{y}) & = & \mathbb{E}\, y_i y_j y_k y_l - \mathbb{E}\, y_i y_j \mathbb{E}\, y_k y_l - \mathbb{E}\, y_i y_k \mathbb{E}\, y_j y_l - \mathbb{E}\, y_i y_l \mathbb{E}\, y_j y_k \end{array} \right. \tag{30}$$

Une définition similaire existe dans le cas général en remplaçant la variable $\mathbf{y}$ par la variable centrée $\mathbf{y} - \mathbb{E}\,\mathbf{y}$.

On voit que dès que les variables y_i, y_j, y_k, y_l sont séparables en deux groupes mutuellement indépendants alors $\mathcal{C}_{ijkl}(\mathbf{y})$ est nul. Les cumulants d'ordre élevé permettent donc de tester l'indépendance au-delà de la simple décorrélation. Par exemple, les cumulants croisés de $\tilde{\mathbf{y}}$ sont tous nuls :

$$\left\{ \begin{array}{lcl} \mathcal{C}_{ij}(\tilde{\mathbf{y}}) & = & \sigma_i^2 \delta_{ij} \\ \mathcal{C}_{ijkl}(\tilde{\mathbf{y}}) & = & k_i \delta_{ijkl} \end{array} \right. \tag{31}$$

où $\sigma_i^2 = \mathbb{E}\,(y_i - \mathbb{E}\, y_i)^2$ est la variance de y_i et $k_i = \mathbb{E}\,(y_i - \mathbb{E}\, y_i)^4 - 3\mathbb{E}^2(y_i - \mathbb{E}\, y_i)^2$ son **kurtosis**.

Le kurtosis constitue une mesure de la gaussianité d'une variable. Elle est nulle pour les variables gaussiennes, positive pour les varibles dites sur-gaussiennes, et négative pour les sub-gaussiennes [Oja01]. Les variables sur-gaussiennes ont typiquement une densité de probabilité avec un "pic" et des "ailes lourdes", c'est-à-dire une densité relativement élevée près de l'espérance et pour les grandes valeurs de la variable. Un exemple classique est la distribution de Laplace (fig. 28). Au contraire, les variables sub-gaussiennes ont une densité de probabilité "plate", elles sont presque constantes au voisinage de l'espérance, et quasi nulles pour les grandes valeurs de la variable. Un exemple classique est la distribution uniforme (fig. 29).

Il est maintenant possible d'estimer la fonction de contraste $\phi_{MI}(\mathbf{y}) = \mathbf{K}(\mathbf{y}|\tilde{\mathbf{y}})$ à partir des cumulants d'ordre 2 et 4 de $\mathbf{y}$ et $\tilde{\mathbf{y}}$. Une expansion d'Edgeworth (développement limité d'une densité de probabilité de variable aléatoire au voisinage d'une densité gaussienne) donne [Car98b]

$$\phi_{MI}(\mathbf{y}) \simeq \frac{1}{4}\phi_2(\mathbf{y}) + \frac{1}{48}\phi_4(\mathbf{y}), \tag{32}$$

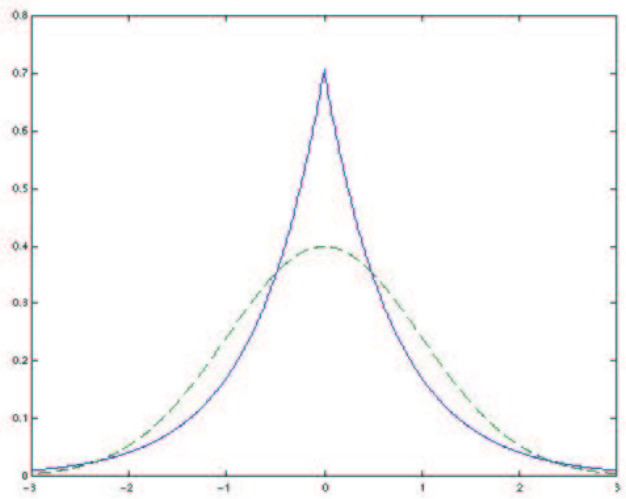

FIG. 28 – Densité de probabilité de Laplace comparée à la densité gaussienne

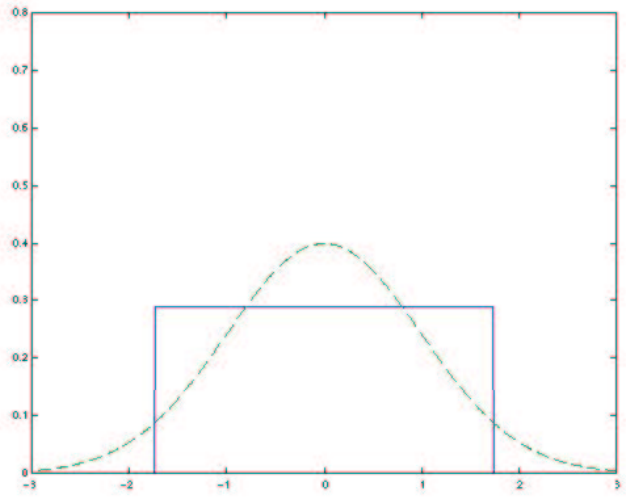

FIG. 29 – Densité de probabilité uniforme comparée à la densité gaussienne

où

$$\phi_2(\mathbf{y}) = \sum_{ij} (\mathcal{C}_{ij}(\mathbf{y}) - \sigma_i^2 \delta_{ij})^2 \quad (33)$$

$$\phi_4(\mathbf{y}) = \sum_{ijkl} (\mathcal{C}_{ijkl}(\mathbf{y}) - k_i \delta_{ijkl})^2 \quad (34)$$

Sous l'hypothèse que les observations comme les sources sont décorrélées et

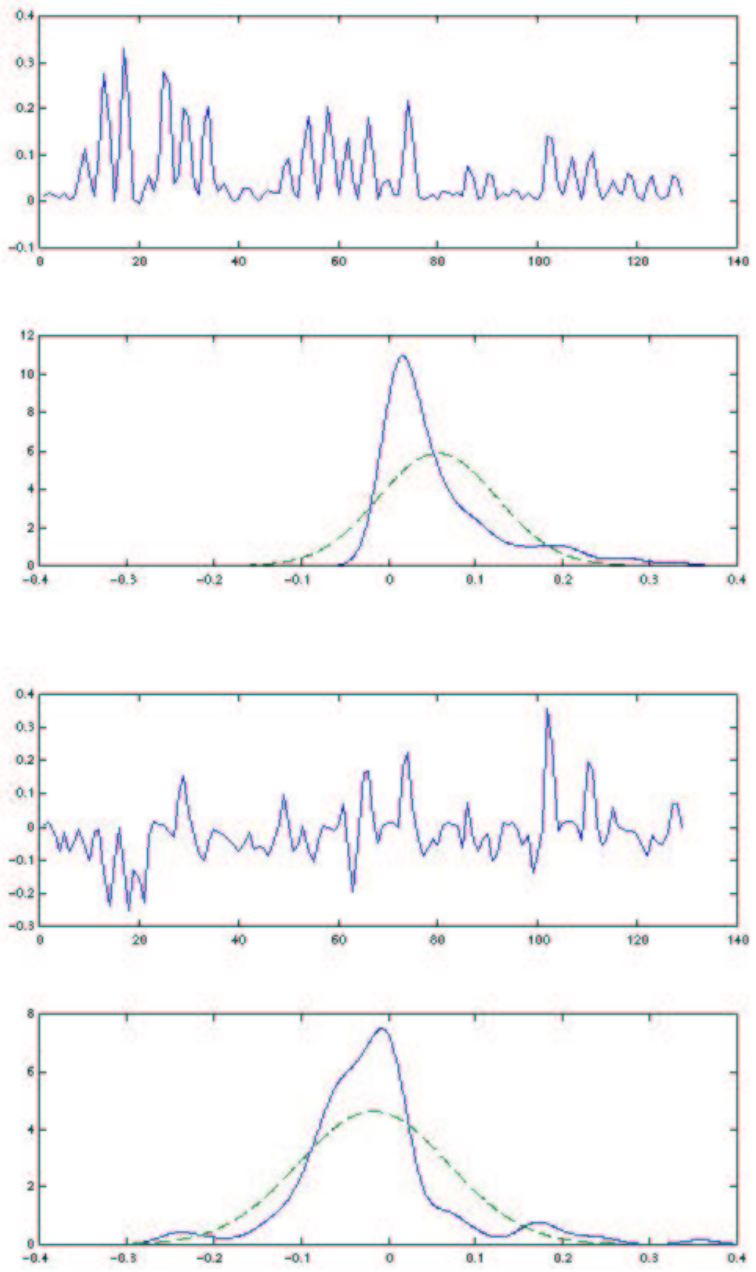

FIG. 30 – Deux spectres "indépendants" résultant d'une ISA sur une somme de signaux périodiques complexes et leurs densités de probabilité (estimées par un noyau gaussien de largeur 0.02)(kurtosis respectifs $0.62.10^{-4}$ et $2.00.10^{-4}$)

de variance unité, le critère de décorrélation $\phi_2(\mathbf{y})$ s'annule, et suite à quelques

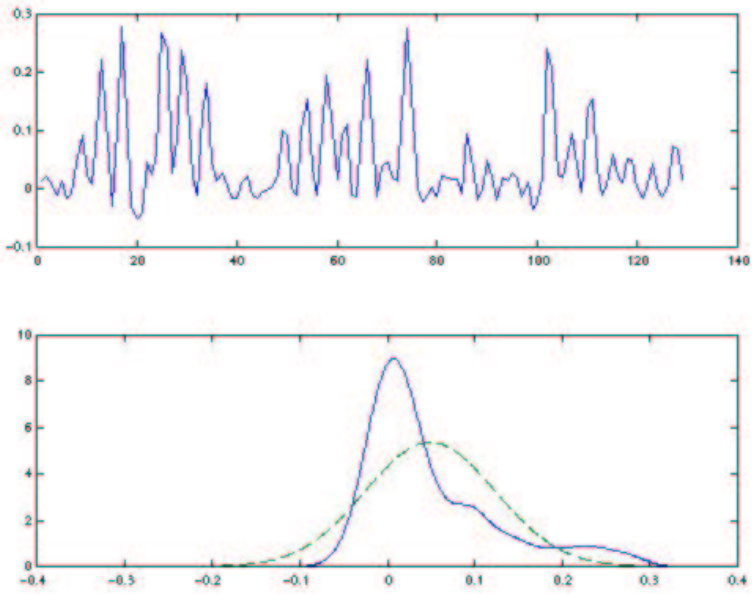

FIG. 31 – Un mélange des deux spectres précédents (poids 0.95 et 0.30) et sa densité de probabilité (kurtosis $0.39.10^{-4}$)

manipulations algébriques on se ramène à la fonction de contraste

$$\phi^0_{ICA}(\mathbf{y}) = -2\sum_i k_i^2 + cst. \tag{35}$$

$\phi^0_{ICA}(\mathbf{y})$ est un critère de non-gaussianité, autrement dit appliquer une ICA revient à maximiser la non-gaussianité des sources estimées. En effet, $\phi^0_{ICA}(\mathbf{y})$ est minimal lorsque les kurtosis k_i ont globalement les valeurs absolues les plus élevées possible. Cette observation très qualitative correspond en fait à la proposition rigoureuse suivante. Considérons une source estimée y_i comme le mélange des sources s_i sous la forme $y_i = \sum_i c_i s_i$, et supposons qu'il existe deux sources i et j telles que leurs kurtosis respectifs vérifient $k_i > 0$ et $k_j < 0$. On sait que la kurtosis de y_i vaut $\mathcal{C}_{iiii}(\mathbf{y}) = \sum_i c_i^4 k_i$. Alors, les points extrémaux de $\mathcal{C}_{iiii}(\mathbf{y})$ en fonction des poids c_i et sous la contrainte $\sum_i c_i^2 = 1$ sont atteints en des points où le vecteur des poids n'a qu'une composante non nulle, autrement dit aux points où y_i est égal à l'une des sources s_i [Del95].

On peut comprendre cette idée de maximiser la non-gaussianité en remarquant que le mélange normalisé de deux variables indépendantes a souvent une densité de probabilité plus gaussienne que celles des variables de départ, en prenant la valeur absolue du kurtosis comme mesure de non-gaussianité. On peut en voir un exemple dans le mélange de la figure 31 á partir des spectres de la figure 30. La densité de probabilité du mélange a un "pic" plus faible près

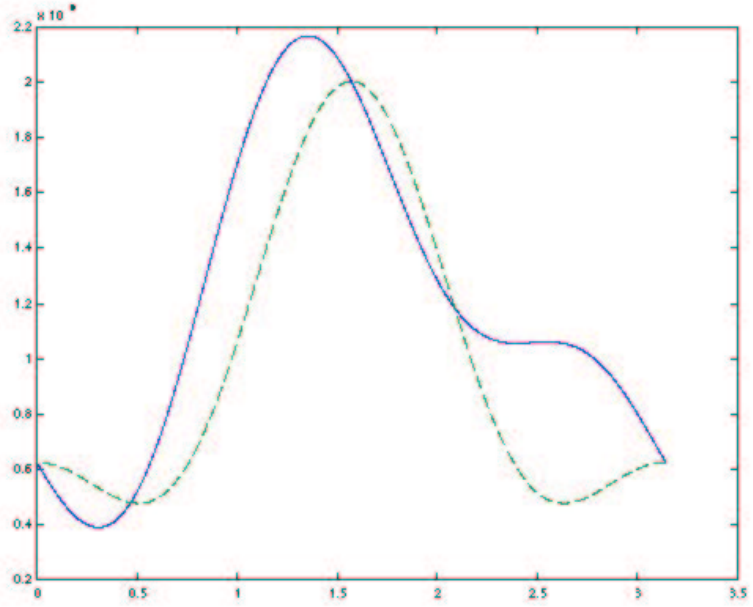

FIG. 32 – Tracé du kurtosis du signal en fonction de l'angle de mélange des deux composantes de la figure 30 (le minimum correspond au tracé de la figure 31)(en pointillé, tracé théorique pour deux composantes réellement indépendantes)

de zéro, et les valeurs moyennes de la variable sont atteintes plus souvent que dans les composantes non mélangées. Cependant, le tracé du kurtosis montre que ce n'est pas le cas pour tous les mélanges, particulièrement dans le cas où les variables ne sont pas réllement indépendantes (fig. 32).

On peut maintenant fournir une interprétation non statistique de l'ICA en remarquant que les observations de variables sur-gaussiennes sont à distribution "creuse" (**sparse**) : elles ont un support réduit (où on entend par support celui des observations significativement différentes de l'espérance, pour rester cohérent dans le cas de variables non centrées). Cela se comprend dans la mesure où le "pic" au voisinage de l'espérance correspond à un grand nombre d'observations au voisinage de l'espérance, alors que les "queues lourdes" indiquent un nombre non négligeable d'observations très différentes.

Dans le formalisme de l'ISA sur le spectrogramme, les composantes que l'on veut estimer sont des spectres, avec souvent quelques valeurs élevées (les résonances) et beaucoup de valeurs quasi nulles quand il s'agit de signaux périodiques peu bruités. Si on les considère comme des variables aléatoires, elles sont donc sur-gaussiennes (fig. 30). Les observations, étant constituées de mélanges des sources, ont des distributions plus gaussiennes, donc avec des supports plus larges. L'ICA tend à redonner aux composantes leur caractère

sur-gaussien, et donc à les rendre les plus "sparse" possible.
On peut aussi voir cette propriété sous l'angle de la réduction de la redondance entre les composantes : dans le cadre de la théorie de l'information c'est exactement le sens du principe de minimisation de l'information mutuelle [Sma01].
Cette propriété se voit très clairement en comparant les figures 7 et 9.

6 Difficultés du modèle : améliorations possibles et autres modèles

Nous avons été confrontés à plusieurs questions ou problèmes lors de l'application du modèle de l'ISA sur le spectrogramme en module. Nous avons imaginé plusieurs réponses possibles à ces problèmes que nous exposons dans le cas de l'extraction des 3 percussions dans l'extrait de percussions de la partie 4.3.

6.1 Utilisation du module du spectrogramme - Essais avec le spectrogramme au carré ou avec une représentation temps-échelle

On peut s'interroger sur la motivation du choix du module du spectrogramme comme TFD pour rechercher une base de l'espace fréquentiel. En fait, le seul critère de choix d'une TFD convenant au modèle d'ISA est que cette TFD donne une signification énergétique aux composantes extraites par la SVD. Par exemple, dans le cas du spectrogramme en module, la somme des coefficients au carré d'une colonne ou d'une ligne représente exactement l'énergie du signal à cet instant ou à cette fréquence dans la fenêtre d'analyse considérée. Ainsi, comme on l'a expliqué dans la partie 3.3, effectuer la SVD du spectrogramme en module $S = \sum_{i=1}^{min(n,m)} \sigma_{ii} u_i v_i^T$ a un sens : u_i est un spectre en amplitude, v_i une amplitude temporelle, et σ_{ii} la "proportion" en amplitude de u_i et v_i dans le signal.

Remarquons que les u_i comme les v_i ont des coefficients négatifs. Les composantes $\sigma_{ii} u_i v_i^T$ ont donc des coefficients négatifs aussi. Le sens à donner à ces coefficients se voit dans la formule de reconstitution ci-dessus : lorsqu'on additionne toutes les composantes, on retrouve la TFD initiale. Les coefficients négatifs de certaines composantes se combinent avec les coefficients positifs des autres pour donner une TFD positive. Ils correspondent à la phase des formes d'onde reconstituées, dans ce sens que deux signaux de même fréquence en opposition de phase s'annulent.

J'ai testé l'algorithme d'ISA avec trois autres TFDs : le spectrogramme complexe, le spectrogramme en module au carré, et le spectrogramme à Q constant en module. Le cas du spectrogramme complexe est abordé dans la partie 6.2.

La SVD du spectrogramme en module au carré a donne à u_i, v_i et σ_{ii} le sens suivant : u_i est un spectre en énergie, v_i une variation temporelle en énergie, et σ_{ii} la "proportion" en énergie de u_i et v_i dans le signal. Les résultats sont présentés dans la figure 33, où on a gardé quatre composantes pour 86.6% de l'énergie.
L'utilisation du spectrogramme en module au carré a renforcé les composantes de fort volume sonore : les composantes 1, 2 et 3 représentent la "bass drum", et la composante 4 la "snare drum". Le "hi-hat" n'apparaît que de façon négligeable dans les composantes 1 et 4. On peut le comprendre sur la figure

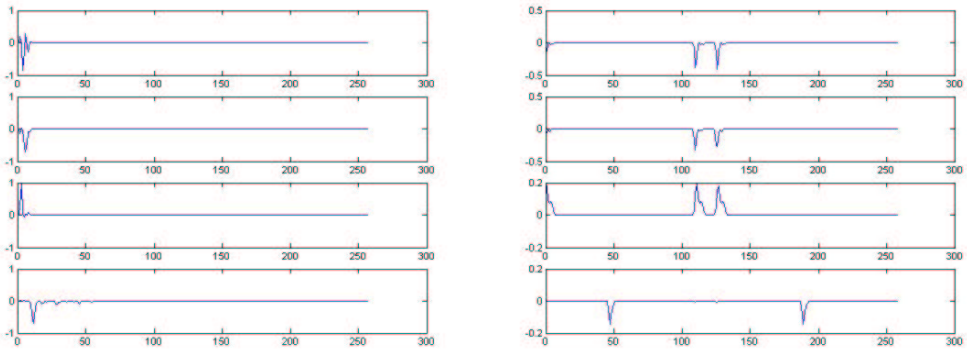

FIG. 33 – Les caractéristiques spectrales et les poids temporels après ICA (spectrogramme en module au carré de l'extrait de percussions)

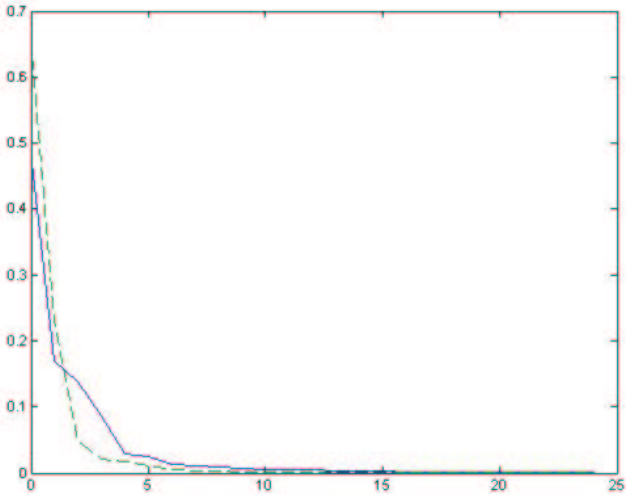

FIG. 34 – Les proportions en énergie des 25 premières composantes issues de la SVD du spectrogramme en module au carré de l'extrait de percussions (en pointillé celles du spectrogramme en module)

34 : à part ses deux premières composantes, la SVD sur le spectrogramme en module au carré a des composantes d'énergies plus élevées que celles de la SVD sur le spectrogramme en module. Il faut donc plus de composantes pour rendre compte de toute l'énergie du signal.
Ce n'est donc pas une bonne idée d'utiliser le spectrogramme en module au carré dans l'ISA pour une extraction directe des sources, par contre comme on détecte prioritairement les composantes de fort volume sonore, on pourrait s'en servir dans une approche de séparation par récurrence, qui consisterait à extraire par ISA la composante de plus fort volume sonore, à la soustraire du

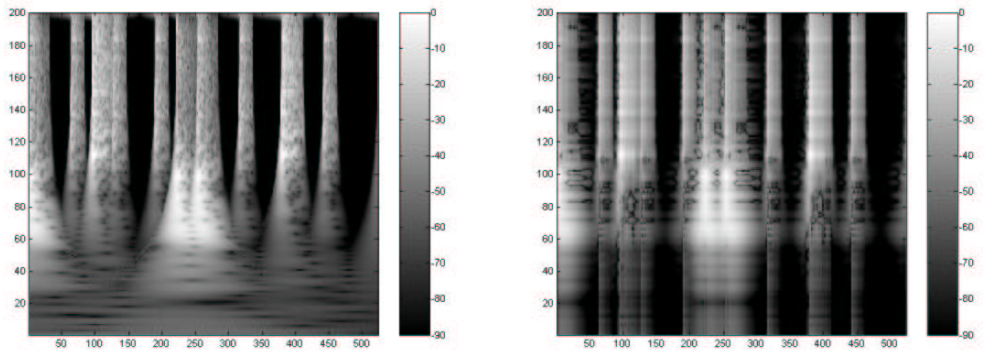

FIG. 35 – Spectrogramme à Q constant de l'extrait de percussions, avant et après SVD

spectrogramme et à recommencer jusqu'à avoir extrait suffisamment de composantes. J'ai effectué des tests de cette méthode, mais malheureusement les résultats ne sont pas satisfaisants, surtout parce que la composante de plus fort volume sonore n'est pas extraite parfaitement et contient un peu des autres instruments.

La SVD du spectrogramme à Q constant en module (fig. 35) a la même signification que celle du spectrogramme en module, lorsqu'on prend comme TFD la transformée en ondelettes renormalisée, définie pour un signal continu par :

$$S(a,b) = \frac{1}{a}\int_{-\infty}^{+\infty} s(t)\psi^*\left(\frac{t-b}{a}\right)\,dt, \tag{36}$$

où ψ est l'ondelette mère, et qui vérifie :

$$\|s\|^2 = cste(\psi)\int_0^{+\infty}\int_{-\infty}^{+\infty} |S(a,b)|^2\,db\,dlog(a) \tag{37}$$

Généralement, le spectrogramme à Q constant est utile pour visualiser un son car l'information utile couvre toute la TFD, contrairement au spectrogramme où elle est souvent concentrée en basse fréquence. On peut donc prendre moins de points fréquentiels que dans le calcul du spectrogramme, et réduire ainsi le temps de calcul.
J'ai écrit un programme qui effectue le calcul en choisissant une ondelette de Gabor de largeur variant de 3 à environ 3000 échantillons, et 200 points fréquentiels. On garde quatre composantes pour 94.8% de l'énergie. Les résultats sont présentés sur la figure 36.
Malgré la part importante d'énergie conservée par la SVD, seule la "snare drum" est à peu près extraite dans la composante 4. Les autres composantes sont des mélanges des trois instruments. Cela peut se comprendre ainsi : sur la figure 35, on voit que la forme dessinée par la "bass drum" en basses fréquences est arrondie, ce qui veut dire que son spectre change beaucoup au cours du temps.

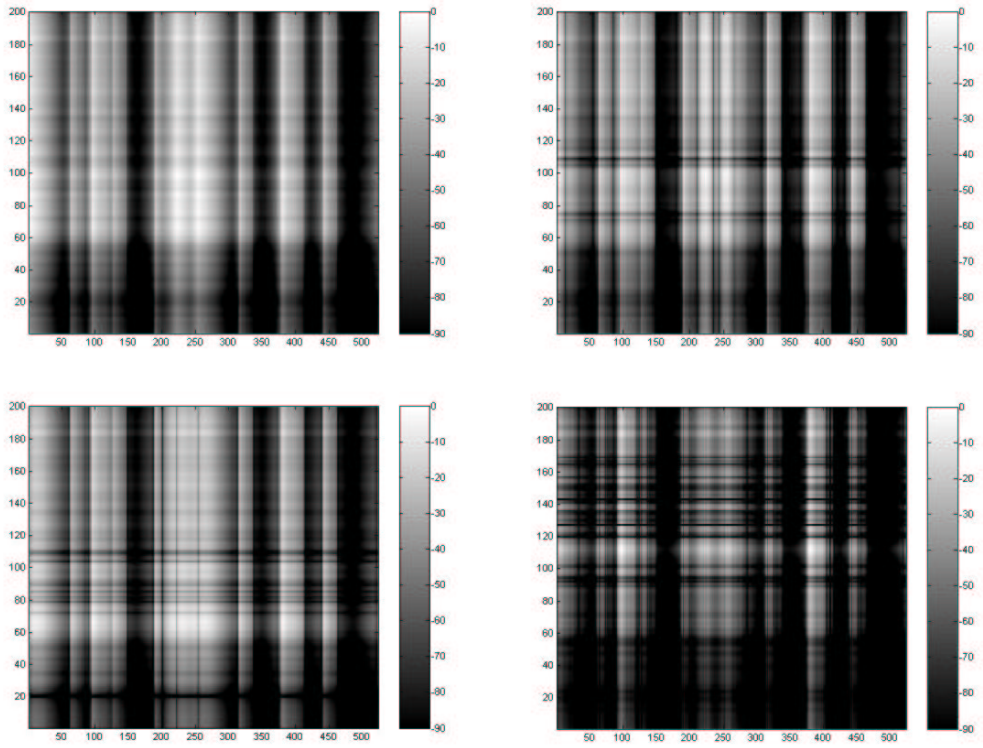

FIG. 36 – Les quatre TFDs indépendantes séparées du spectrogramme à Q constant de l'extrait de percussions

Comme les composantes estimees par l'ISA ont un spectre constant, la "bass drum" est forcément répartie sur plusieurs composantes. La "snare drum", dont la forme sur le spectrogramme est moins arrondie à cause de sa fréquence plus élevée, ne souffre pas de ce problème.
L'ISA sur le spectrogramme à Q constant n'est donc pas non plus applicable directement : sa bonne résolution fréquentielle en basses fréquences ne compense pas le défaut d'une mauvaise résolution temporelle. Par contre, on pourrait penser à l'appliquer dans le modèle non stationnaire de l'ISA [Del01].

6.2 Qualité de l'inversion de la TFD - Essais avec le spectrogramme complexe

Un des gros problèmes que nous rencontrons avec le modèle d'ISA concerne l'inversion du spectrogramme pour retrouver les formes d'ondes correspondant aux composantes. En effet, la méthode de Griffin et Lim, même précédée d'une estimation de la phase, reste coûteuse en calculs et donne des résultats bruités. J'ai donc appliqué le modèle d'ISA au spectrogramme complexe, en espérant obtenir des composantes avec une phase cohérente, ou du moins pas trop éloignée de celle d'un spectrogramme.

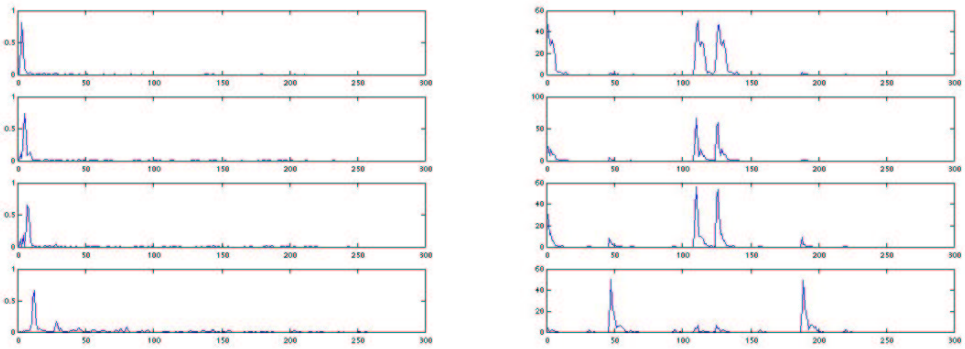

FIG. 37 – Les modules des caractéristiques spectrales et des poids temporels après ICA (spectrogramme complexe de l'extrait de percussions)

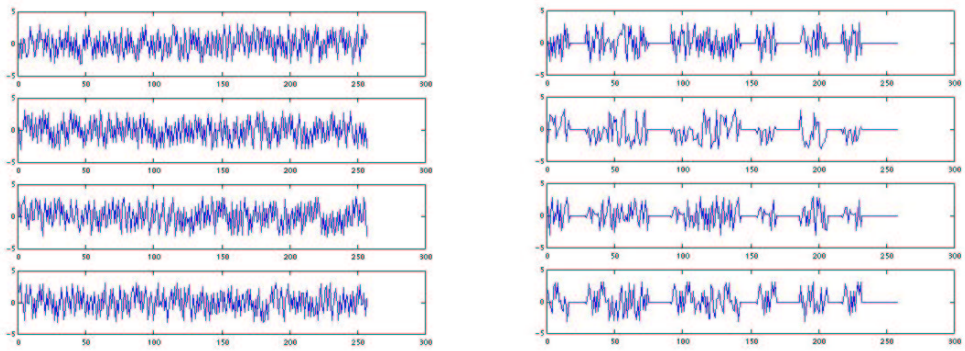

FIG. 38 – Les phases des caractéristiques spectrales et des poids temporels après ICA (spectrogramme complexe de l'extrait de percussions)

La SVD a le même sens que celle du spectrogramme en module, à condition de remplacer la transposition ^{T} par la transposition-conjugaison *. Quant à la fonction de contraste ϕ_{MI}, si elle reste valable en considérant des densités de probabilité de variables complexes, on ne peut plus appliquer la méthode de gradient définie ci-dessus. J'ai donc utilisé une méthode par diagonalisation conjointe de matrices de cumulants [Car99] implémentée dans l'algorithme JADE [JADE] de Jean-François Cardoso que j'ai modifié pour accepter des variables d'espérance non nulle. On garde quatre composantes pour 72.5% de l'énergie du signal. Les résultats sont présentés dans la figure 37.

Comme dans le cas du spectrogramme en module au carré, on remarque que les composantes de fort volume sonore ont encore été extraites en priorité. La SVD (dont on peut voir les premiers coefficients sur la figure 39) a des compo-

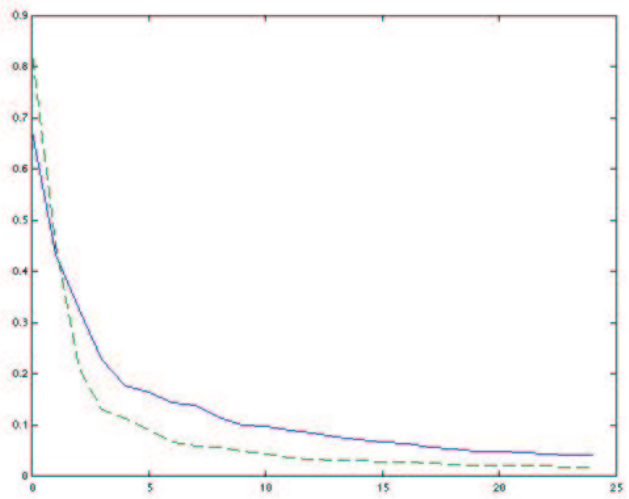

FIG. 39 – Les proportions en amplitude des 25 premières composantes issues de la SVD du spectrogramme complexe de l'extrait de percussions (en pointillé celles du spectrogramme en module)

santes d'énergies plus fortes que celles du spectrogramme en module (sauf les deux premières). En effet, les coefficients des u_i et des v_i présentent des variations de phase très importantes (fig. 38) qui ont tendance à s'annuler dans la reconstitution $\sigma_{ii}u_iv_i^*$, donnant des composantes d'énergies relativement élevées comparées à l'énergie totale. De plus, on remarque sur la figure 37 que le modèle opère des confusions entre les instruments, du fait encore de ces variations de phase désordonnées qui perturbent la fonction de contraste. La reconstitution des formes d'onde n'est pas non plus facilitée pour cette même raison.

La solution serait donc de modifier le spectrogramme complexe pour qu'il possède des variations de phase moins importantes, par exemple presque nulles pour une fréquence donnée sauf aux moments où le signal lui-même varie beaucoup comme les attaques. J'ai donc effectué des essais en remplaçant la phase par sa dérivée temporelle, on obtient cette fois 86.9% de l'énergie avec quatre composantes, mais il reste encore des confusions entre "bass drum" et "snare drum" à l'extraction : la phase ne semble pas apporter une information utile à cette fonction de contraste. Peut-être pourrait-on s'en servir pour une autre fonction de contraste qui s'ajouterait à ϕ_{MI} calculée sur le module. Ou calculer beaucoup de composantes sur le spectrogramme en module et se servir de la phase dans l'étape de regroupement alors nécessaire.

6.3 Extraction des composantes de faible volume sonore - Renormalisation du spectre

Les autres TFDs qu'on a utilisé ci-dessus dans le cadre de l'ISA n'ont pas donné de résultats satisfaisants concernant l'extraction des composantes de faible volume sonore. Une approche pour augmenter leur volume et mieux les extraire consiste alors à changer les énergies des canaux fréquentiels du spectrogramme en module [Cas98]. Par exemple, dans le cas de l'extrait de percussions, la séparation du "hi-hat" est gênée par le manque de hautes fréquences dans le spectre. En augmentant ces fréquences, on devrait mieux séparer le "hi-hat" sans trop modifier l'extraction de la "bass drum", puisqu'elle ne couvre que les basses fréquences.

J'ai utilisé une approche sans a priori sur les composantes à extraire qui consiste à calculer l'énergie E_i de chaque canal fréquentiel (en additionnant les coefficients au carré de la ligne correspondante du spectrogramme), et à les normaliser par une fonction $1/f(E_i)$. On applique ensuite l'algoritme d'ISA pour obtenir des caractéristiques fréquentielles u'_i dont on renormalise les coefficients par $f(E_i)$ pour donner les véritables caractéristiques fréquentielles u_i. La renormalisation affecte les u_i, qui ne sont alors plus vraiment indépendants, ni même orthonormés. Après les avoir normés, on projette alors le spectrogramme sur chaque u_i pour trouver les poids temporels v_i correspondants.

La normalisation utilisée par [Cas98] consiste à prendre $f(E_i) = \sqrt{E_i}$, c'est-à-dire à donner la même énergie à tous les canaux fréquentiels. Cette idée peut se révéler utile dans l'analyse de certains bruits (par exemple elle diminuerait l'importance de l'attaque à basse fréquence du bris de verre de la partie 4.1 pour faire mieux ressortir les résonances à haute fréquence), elle ne fonctionne pas lorsqu'on veut extraire des instruments au sens classique du terme car elle a tendance à masquer les informations importantes des spectres comme les résonances. Appliquée avec quatre composantes sur l'extrait de percussions, elle fournit une composante de "bass drum" et une de "snare drum" pas très bien extraites, et deux composantes mélangeant tous les instruments. J'ai essayé d'autres normalisations de type $f(E_i) = E_i^\alpha$, avec α proche de 0 ou de 1, mais on obtient toujours des mélanges des trois instruments comme composantes, et si les résultats sont meilleurs parfois, il semble qu'on ne puisse pas dégager de règle générale.

Par contre, j'ai obtenu des résultats satisfaisants concernant l'extraction du "hi-hat" avec une approche similaire qui consiste à mettre à zéro tous les canaux fréquentiels d'énergie trop élevée. On peut voir sur la figure 40 les résultats de cette approche, en mettant à zéro les canaux dont l'énergie est supérieure à 0.08 fois l'énergie maximale d'un canal. La "bass drum" n'apparaît plus, et la "snare drum" et le "hi-hat" sont plus ou moins bien extraits. On note que la quatrième composante sépare presque le "hi-hat", et que les valeurs du poids temporel correspondant à la "snare drum" sont de signe différent de celui du "hi-hat", mais malheureusement cet exemple ne généralise pas.

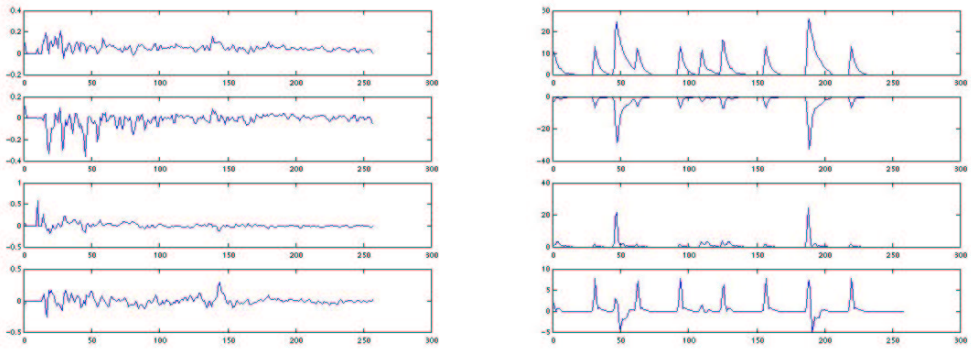

FIG. 40 – Les caractéristiques spectrales et les poids temporels après ICA (spectrogramme de l'extrait de percussions avec canaux fréquentiels de grande énergie mis à zéro)

Il semble donc difficile d'extraire les composantes de faible volume sonore sans les connaître mieux. Nous abordons ce point dans la sous-partie suivante.

6.4 Amélioration de l'extraction d'une composante par ajout de connaissances a priori

Souvent, les sons qu'on veut séparer sont plus ou moins connus. On peut alors utiliser une approche a priori pour mieux les extraire. On examine ici deux hypothèses : extraire le spectre d'un son quand on connaît les instants où il est joué et vice-versa.

L'approche connaissant les instants où joue l'instrument est adaptée à un instrument qui ne joue pas en continu (comme souvent les percussions), mais peut s'appliquer en étendant la notion d'instrument à une note donnée d'un instrument donné. Par exemple, dans l'extrait de percussions, le "hi-hat" joue à huit instants bien précis (fig. 41). J'ai donc mis des silences aux endroits où il ne jouait pas pour augmenter son importance dans le signal. En pratique, on peut savoir ces endroits sur une partition, ou à l'écoute, tout en connaissant le temps de décroissance de sa forme d'onde. Le nouveau spectrogramme est présenté dans la figure 42, et les résultats de l'ISA dans la figure 43.
En comparant les poids temporels aux endroits où on sait que le "hi-hat" joue, on voit que la composante correspondant au "hi-hat" est la deuxième, et on en déduit son spectre. Bien plus, en comparant les poids temporels des composantes entre eux, on peut savoir les endroits où le "hi-hat" joue seul, et ainsi récupérer sa forme d'onde exacte. On l'aura alors extrait sans aucune erreur.

On peut reprocher à cette technique son peu d'intérêt : si on connaît les endroits où joue le "hi-hat", on connaît aussi généralement les endroits où il joue

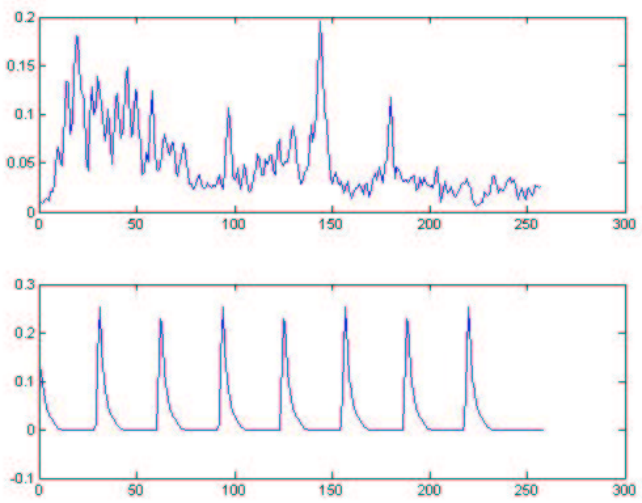

FIG. 41 – Les composantes principales du spectre et de la variation temporelle du "hi-hat" (résultats d'une SVD sur le spectrogramme en module du "hi-hat" extrait à la main)(94.3 % de l'énergie du "hi-hat")

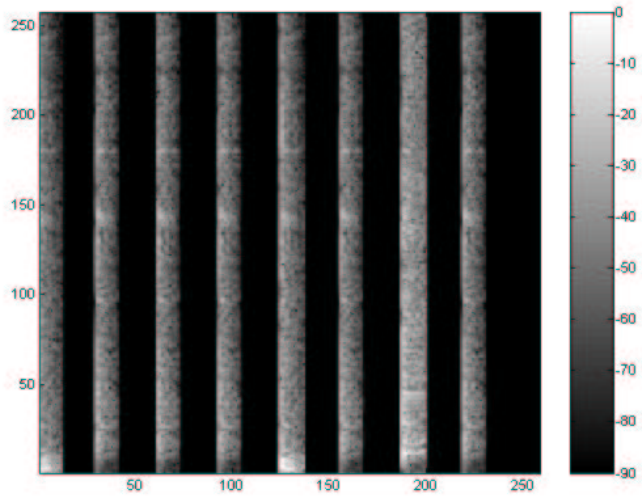

FIG. 42 – Le spectrogramme de l'extrait de percussions rendu silencieux aux endroits où le "hi-hat" ne joue pas

seul, et on a vite fait de récupérer la forme d'onde à ces endroits. Cependant, même lorsque l'instrument ne joue jamais seul, on peut quand même récupérer son spectre par cette technique. J'ai fait des essais en faisant jouer la "bass drum" sur les quatre premières occurences du "hi-hat", et la "snare drum" sur

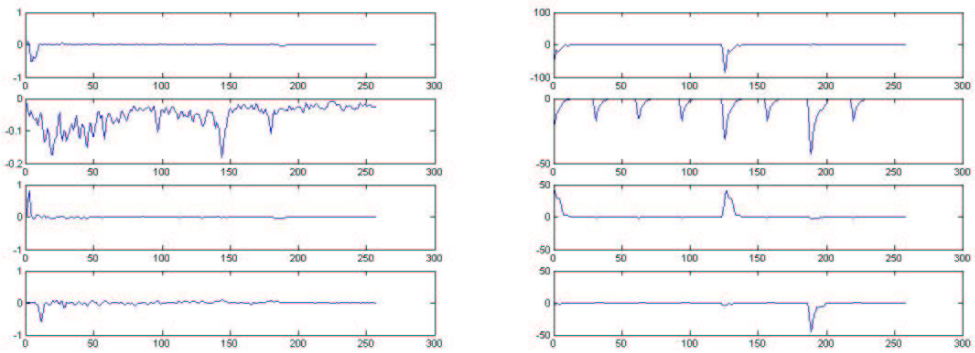

FIG. 43 – Les caractéristiques fréquentielles et les poids temporels correspondants après ICA (spectrogramme de la figure 42)

les quatre dernières, et on obtient des résultats similaires à ceux de la figure 43. En prenant plus de composantes, et en comparant toujours leurs variations temporelles à celles qu'on a supposé, on peut même arriver à reconstituer une grande partie du spectrogramme du "hi-hat". Évidemment, cela ne peut pas marcher si le "hi-hat" joue toujours en même temps que la "bass drum" par exemple. Les deux événements auditifs sont alors associés, et il est normal que nous ne puissions pas les séparer.
Le principal reproche à faire à cette technique est que sur un extrait d'instruments réels, les notes jouées ne sont jamais identiques. Ici, on extrairait non seulement des notes identiques, mais de plus l'extraction serait perturbée du fait qu'elles ne le soient pas en réalité.

L'approche connaissant le spectre approximatif de l'instrument peut s'appliquer dans tous les cas, mais elle est plus difficile à mettre en œuvre. Connaissant le spectre du "hi-hat" (fig. 41), on peut penser qu'une simple projection du spectrogramme sur ce spectre (fig. 44) donne immédiatement les occurrences de l'instrument, mais ce n'est pas le cas puisqu'on détecte encore une fois des occurrences de tous les instruments, et principalement de la "snare drum". Cela signifie qu'une ISA qui trouverait ce spectre pourtant proche de la réalité échouerait à trouver la bonne variation temporelle. Une autre projection montre que le spectre correspondant à la variation temporelle du "hi-hat" (fig. 41) a beaucoup trop de basses fréquences par rapport à la réalité (fig. 44). On ne peut donc pas extraire le "hi-hat" en une seule composante.

J'ai pensé à rajouter avec un poids approprié une autre fonction de coût à la fonction de contraste ϕ_{MI}, de sorte que la fonction totale soit plus faible lorsque les spectres des composantes ressemblent (au sens d'une distance quadratique entre leurs valeurs absolues par exemple) à celui de la figure 41. Mais on obtient des composantes dont les spectres lui ressemblent tous plus ou moins, et qui par conséquent ne séparent pas du tout les instruments. On peut alors penser

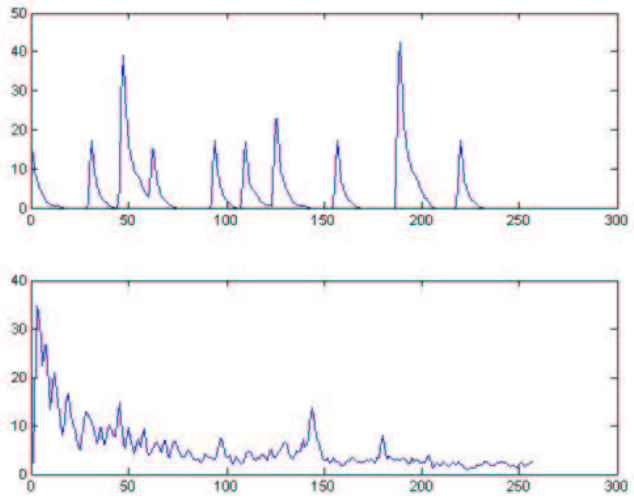

FIG. 44 – Amplitude temporelle correspondant au spectre de la figure 41 et spectre correspondant à l'amplitude temporelle de la figure 41 dans l'extrait de percussions

à modifier la fonction de coût rajoutée en la définissant par exemple de sorte qu'elle ait une valeur faible dès que deux composantes ont un spectre proche de celui de la figure 41, et toutes les autres un spectre très différent. Une telle fonction reste difficile à définir effectivement, de plus elle n'admet pas de gradient calculable et sa minimisation repose donc sur un algorithme de type recuit simulé très lourd en temps de calcul vu le nombre de paramètres.

J'ai donc tenté une approche similaire à celle utilisant les instants où joue l'instrument, c'est à dire en rendant silencieuses certaines parties du spectrogramme. Comme le spectre du "hi-hat" contient de toutes les fréquences, j'ai effectué un seuillage du spectre de la figure 41 en mettant à zéro les fréquences dont l'amplitude ne dépasse pas 0.6 fois l'amplitude maximale, et à un toutes les autres. Cela donne un filtre qu'on applique au spectrogramme, dont tous les canaux fréquentiels ont été normalisés, (fig. 45) avant de faire une ISA. Comme le filtrage supprime beaucoup de fréquences, la normalisation de tous les canaux fréquentiels est nécessaire pour éviter que la faible quantité d'information que contiennent maintenant les caractéristiques fréquentielles soit concentrée dans les valeurs élevées de la "snare drum" en basses fréquences.

Les résultats , présentés dans la figure 46, sont encourageants. L'ISA a effectivement plus ou moins bien séparé les occurences temporelles des deux instruments, en détectant la résonance à haute fréquence du "hi-hat", absente dans la "snare drum". Le "hi-hat" semble être la composante 2, mais il est présent un peu aussi dans les autres composantes, et il est difficile de faire le choix de façon calculatoire. Il faudrait pour cela comparer les spectres des composantes

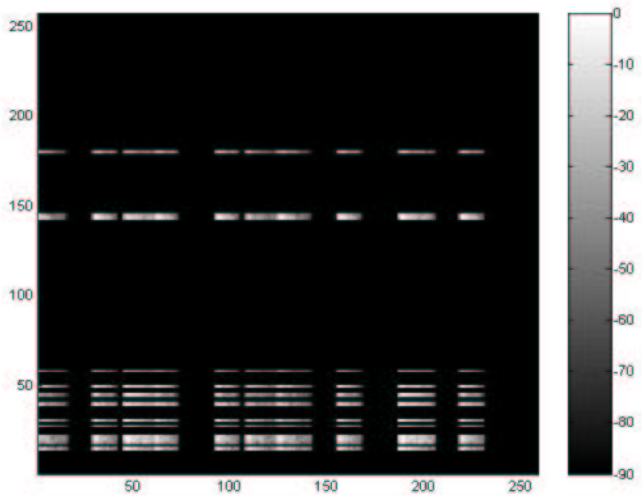

FIG. 45 – Le spectrogramme de l'extrait de percussions normalisé puis rendu silencieux aux endroits où le spectre du "hi-hat" est trop faible

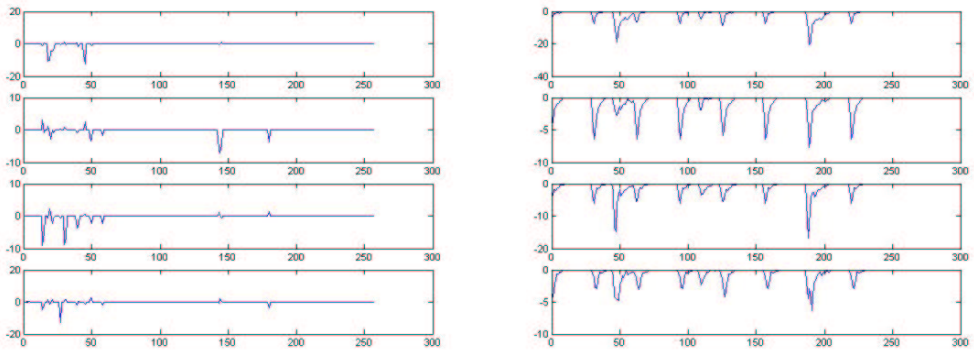

FIG. 46 – Les caractéristiques fréquentielles et les poids temporels correspondants après ICA sur les caractéristiques fréquentielles (spectrogramme de la figure 45)

extraites avec celui du "hi-hat", mais le fait d'avoir dû seuiller ce spectre complique beaucoup la tâche.
Remarquons que cette méthode est un peu plus fiable et donne quelques informations de plus qu'un simple filtrage suivi de la détection des maxima du signal filtré. Elle fournit en effet une meilleure estimation du volume sonore de l'instrument détecté et de la forme de son enveloppe temporelle. Lorsqu'un autre instrument de spectre pas trop différent joue en même temps, le signal filtré a des valeurs non nulles sur les occurences de cet instrument. Au contraire, l'ISA détecte les différences de spectre et sait retrouver des poids temporels plus

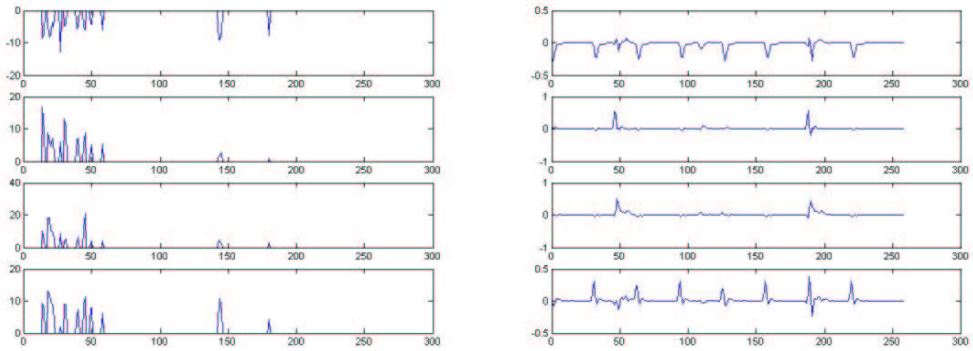

FIG. 47 – Les caractéristiques fréquentielles et les poids temporels correspondants après ICA sur les poids temporels (spectrogramme de la figure 45)

proches de la réalité.

J'ai eu des résultats encore plus encourageants en utilisant le même spectrogramme avec les canaux fréquentiels normalisés puis filtrés, à ceci près que l'ICA a été appliquée sur les poids temporels et non pas sur les caractéristiques fréquentielles. Les résultats sont présentés dans la figure 47. Grâce à sa propriété de rendre les signaux plus "sparse", l'ICA a séparé les deux occurrences de la "snare drum" des occurences du "hi-hat", mais encore une fois il reste difficile de choisir automatiquement quelle est la bonne composante.

Remarquons enfin qu'il est possible de combiner les deux approches a priori de cette sous-partie. Même lorsque le spectre exact de l'instrument n'est pas connu, il est possible d'utiliser l'approche par filtrage en prenant un filtre assez large. Dès que les occurences approximatives de l'instrument sont trouvées, on peut utiliser cette nouvelle connaissance pour appliquer la méthode avec a priori sur la localisation temporelle des instruments, et trouver alors le spectre et les occurences presque exacts de l'instrument.

7 Conclusion et perspectives

Cette étude propose des éléments de réponse au problème de la séparation de sources sur des signaux monophoniques.

Le modèle d'ISA sur le spectrogramme se révèle particulièrement adapté à l'analyse de bruits ou à l'extraction d'attaques. Cependant, il est insuffisant pour extraire les sources d'un mélange quelconque de plusieurs instruments, surtout celles de faible volume sonore. L'utilisation du spectrogramme complexe ou du spectrogramme à Q constant n'apporte pas de solution satisfaisante. Par contre, la prise en compte d'informations a priori sur les sources, comme leur spectre approximatif ou leur localisation temporelle, donne des résultats assez précis permettant d'extraire des sources proches de la réalité.

Cette étude se poursuivra dans deux directions parallèles.

D'une part, on étudiera les applications du modèle à la reconnaissance et la classification d'instruments. On sait que les statistiques d'ordre élevé permettent de distinguer les caractéristiques des spectres [Dub97a][Dub97b], et on peut envisager une généralisation tenant compte des caractéristiques spectrales mais aussi de leur évolution temporelle, ce qui permettrait de classifier facilement les instruments à la fois par leurs attaques et par leurs parties soutenues.

D'autre part, on s'intéressera à la séparation d'instruments musicaux, dans le cadre d'un son monophonique ou stéréophonique.

Pour cela, on cherchera des méthodes permettant d'utiliser des connaissances a priori sur ces instruments pour améliorer l'extraction aussi bien en qualité qu'en rapidité. Intégrer les distributions des sources dans l'approche par maximum de vraisemblance constitue un facteur de difficulté : une erreur, même légère, peut donner un résultat opposé à celui escompté [Car98b]. D'autres algorithmes [Pea96] ont été proposés, souvent dans des cas très particuliers comme les sources de module constant [Gam97] utilisées en communications [Tor98], mais le domaine reste peu étudié. On pourra penser à intégrer nos recherches sur la classification d'instruments dans cette optique. Dans le même ordre d'idées, la prise en compte d'informations spatiales ou de modèles de mixage apporte des connaissances qu'il serait utile d'exploiter.

On ne se limitera pas au modèle d'ISA sur le spectrogramme décrit dans cette étude, en explorant d'autres possibilités comme l'utilisation d'une ICA directement sur des portions de signal [Dub01] ou de spectrogramme, de manière à trouver une base du signal constituée de signaux "creux" [Zib00] ou de représentations temp-fréquence "creuses", ou en essayant de s'aider d'autres techniques d'analyse du signal comme le Matching Pursuit. Ces techniques reposant sur l'étude d'un très grand nombre de composantes, on s'attachera enfin à la phase de regroupement [Del01] de ces composantes par le calcul de distances appropriées.

Références

[Ant01] A. Antoniadis. Univariate density estimation. *École de printemps : de la séparation de sources à l'analyse en composantes indépendantes*, Villard-de-Lans, 2001

[Bel95] A.J. Bell and T.J. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation* 7(6) :1004-1034, 1995

[Bre90] A.S. Bregman. Auditory scene analysis : the perceptual organization of sound. *MIT Press*, 1990

[BSS] Blind Source Separation :
`http ://www.media.mit.edu/~westner/bssd.html`

[Car98a] J.-F. Cardoso. Multidimensional independent component analysis. *Proc. ICASSP98*, Seattle, 1998

[Car98b] J.-F. Cardoso. Blind signal separation : statistical principles. *Proc. IEEE*, 90(8) :2009-20026, oct. 1998

[Car99] J.-F. Cardoso. High order contrasts for independent component analysis. *Neural computation*, 11 :157-192, 1999

[Car01] J.-F. Cardoso. *Communication privée*, 2001

[Cas98] M.-A. Casey. Auditory group theory : with application to statistical basis methods for structured audio. *PhD thesis*, MIT, Media Lab., 1998

[Cas00] M.-A. Casey and A. Westner. Separation of mixed audio sources by independent subspace analysis. *Proc. ICMC2000*, 2000

[Com94] P. Comon. Independent component analysis, a new concept ? *Signal Processing*, Elsevier, 36(3) :287-314, apr. 1994

[Del95] N. Delfosse and P. Loubaton. Adaptative blind separation of independent sources : a deflation approach. *Signal Processing*, 45 :59-83, 1995

[Del01] B. Delezoide. Séparation de sources audio par analyse en sous-espaces indépendants : modèle non stationnaire et algorithme de regroupement. *Rapport de stage de DEA ATIAM*, 2001

[Dub97a] S. Dubnov and X. Rodet. Statistical Modeling of Sound Aperiodicities. *Proc. ICMC97*, Thessaloniki, sept. 1997

[Dub97b] S. Dubnov and N. Tishby. Analysis of sound textures in musical and machine sounds by means of higher order statistical features. *Proc. ICASSP*, Munich, 1997

[Dub01] S. Dubnov. *Communication privée*, 2001

[FastICA] FastICA package for MATLAB :
`http ://www.cis.hut.fi/projects/ica/fastica/`

[Gam97] F. Gamboa and E. Gassiat. Source separation when the input sources are discrete or have constant modulus. *IEEE Trans. on Sig. Proc.*, 45(12) :3062-3072, 1997

[Gri84] D. Griffin and J. Lim. Signal estimation from modified short time Fourier transform. *IEEE Trans. on Acous., Speech and Sig. Proc.*, 32 :236-242, 1984

[Hyv97] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7) :1483-1492, 1997

[Hyv00] A. Hyvärinen and P.O. Hoyer. Emergence of phase and shift invariant features by decomposition of natural images into independent features subspaces. *Neural Computation*, 12(7) :1705-1720, 2000

[ICACentral] ICA Central :
`http ://sig.enst.fr/~cardoso/icacentral/index.html`

[ICA2001] ICA 2001 : Third International Conference on Independent Component Analysis and Signal Separation, San Diego, California, December 9-13, 2001 : `http ://ica2001.ucsd.edu/`

[JADE] JADE for MATLAB :
`ftp ://sig.enst.fr/pub/jfc/Algo/Jade/jade.m`

[Lat96] L. Lathauwer, D. Callaerts, B. Moor and J. Vandrewalle. Fetal electrocardiogram extraction by source subspace separation. *Proc. IEEE SP Workshop on Stat. Sig. Array Proc.*, pp. 356-359, 1996

[Lee98] T. Lee, M. Girolami, A. Bell and T. Sejnowski. A unifying information-theoretic framework for independent component analysis. *Int. Jour. on Math. and Comp. Modeling*, 1998

[MIalgorithm] MIalgorithm :
`http ://helio.lis.inpg.fr :80/weblis1/demo/ICAdemo/download1.html`

[Oja01] E. Oja. ICA and non gaussianity. *École de printemps : de la séparation de sources à l'analyse en composantes indépendantes*, Villard-de-Lans, 2001

[Pea96] B.A. Pearlmutter and L.C. Parra. A context-sensitive generalisation of ICA. *Int. Conf. on Neural Information Proc.*, Hong-Kong, 1996

[Pha92] D.T. Pham, P. Garrat and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. *Proc. EUSIPCO*, pp.771-774, 1992

[Sla94] M. Slaney, D. Naar and R.F. Lyon. Auditory model inversion for sound separation. *Proc. ICASSP94*, 2 :77-80, 1994

[Sma01] P. Smaragdis. Redundancy reduction for computational audition, a unifying approach. *PhD thesis, MIT*, 2001

[Tal97] A. Taleb and C. Jutten. Entropy optimization - Application to blind separation of sources. *ICANN97, Lausanne*, 1997

[Tor98] K. Torkkola. Blind signal separation in communications : making use of known signal distributions. *IEEE DSP Workshop*, Bryce Canyon, aug. 1998

[Zib00] M. Zibulevsky and B.A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Technical Report CS99-1*, CS dept., Univ. of New Mexico.