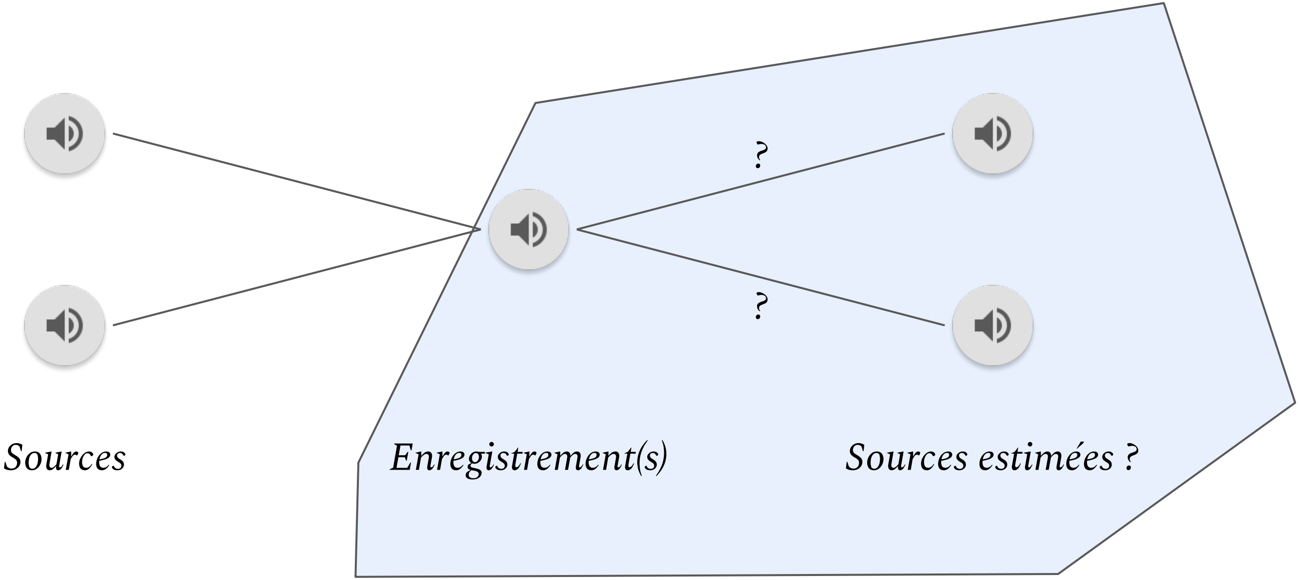
Séparation de sources musicales

Introduction

Introduction

Séparation de sources

But : isoler N sources à partir de M enregistrements



Introduction

Formulation

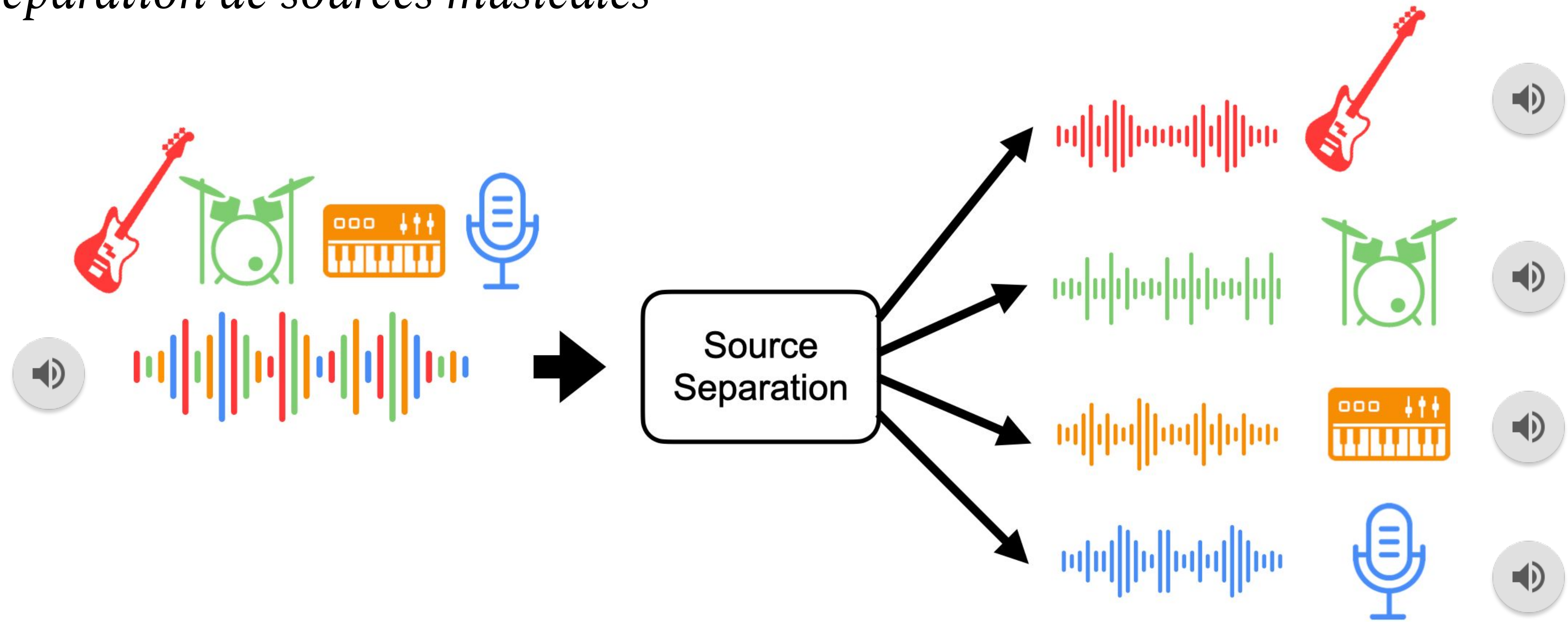
But : isoler N sources à partir de M enregistrements

$$y_i(t) = \sum_{j=1}^{N} a_{ij} s_j(t)$$

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_n \end{pmatrix}$$

Introduction

Séparation de sources musicales



Open Source Tools & Data for Music Source Separation, Ethan Manilow and Prem Seetharman and Justin Salamon

Introduction

Applications

- Directes :
    - Re-mixage
    - Débruitage
- Indirectes :
    - Transcription
    - Changement de tempo, de hauteur
    - Analyse des paroles
    - …

Introduction

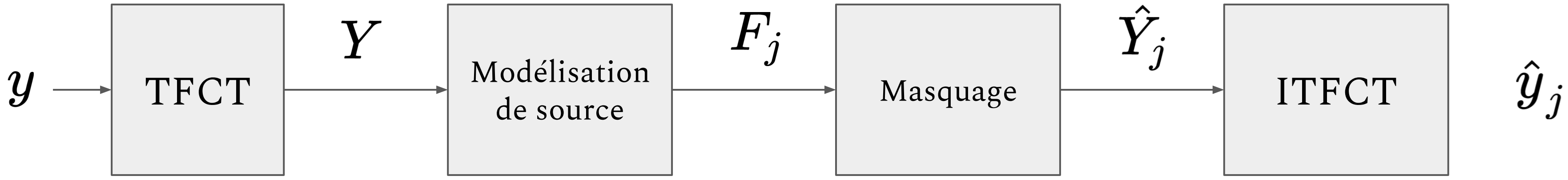
Difficultés

- Sources très corrélées et mêlées
- Souvent, effets non linéaires lors du mixage
- Très difficile de faire une séparation de qualité acceptable

Bases de la séparation de source

Bases de la séparation de source

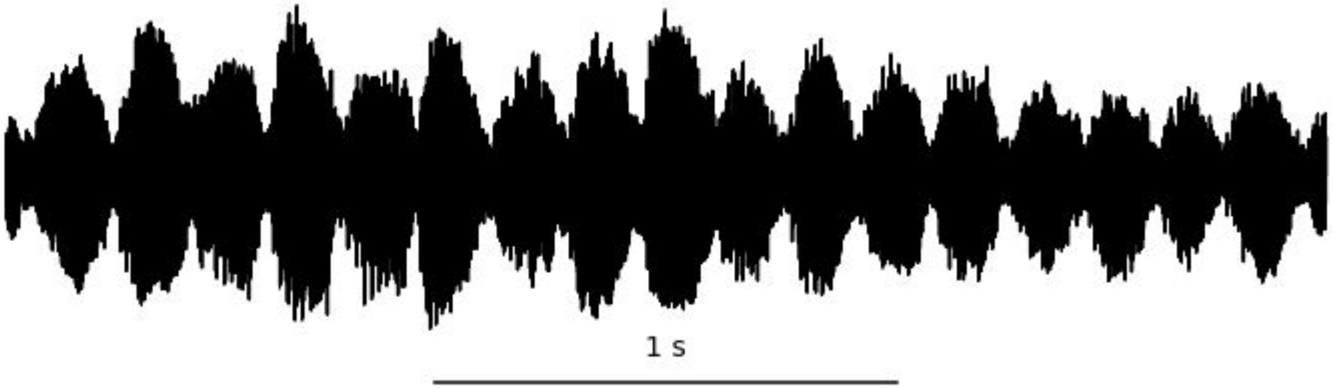
Schéma classique



Bases de la séparation de source

Représentations d’un signal sonore

Formes d’ondes

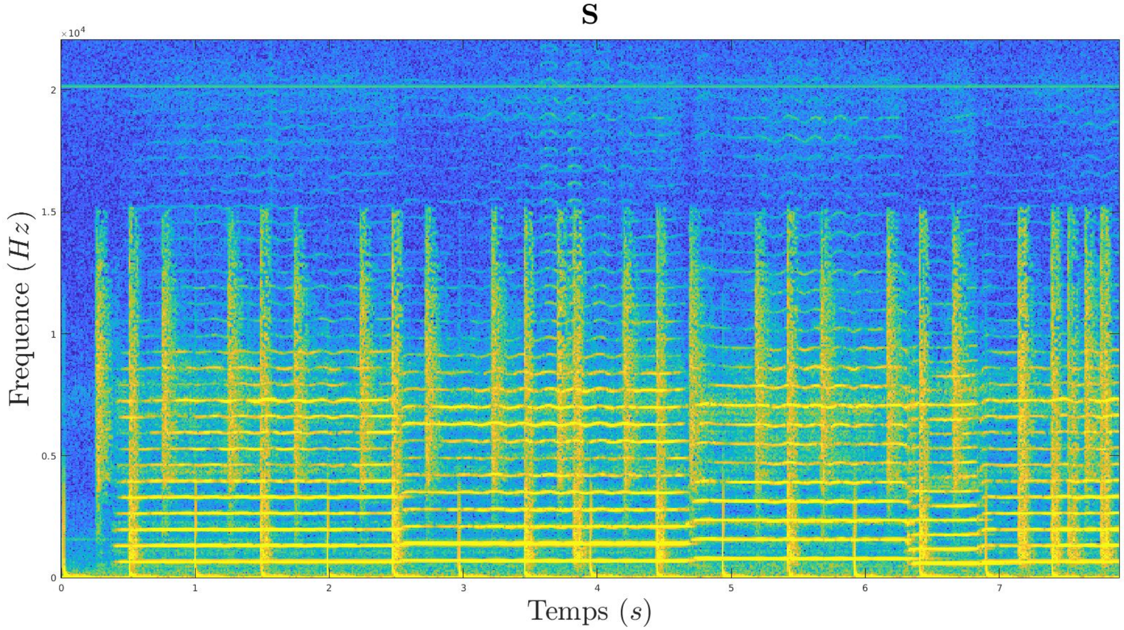


Open Source Tools & Data for Music Source Separation, Ethan Manilow and Prem Seetharman and Justin Salamon

Bases de la séparation de source

Représentations d'un signal sonore

Représentations Temps-Fréquence



Bases de la séparation de source

Masquage

$$
\begin{aligned}
y &= x_1 + x_2 \\
Y &= X_1 + X_2 \\
|Y| &= |X_1 + X_2| \\
&\neq |X_1| + |X_2|
\end{aligned}
$$

Bases de la séparation de source

Masquage binaire

$$M_1 = (F_1 > F_2)$$
$$M_2 = (F_2 > F_1)$$

puis

$$\hat{X}_1 = M_1 \odot Y$$
$$\hat{X}_2 = M_2 \odot Y$$

Bases de la séparation de source

Masquage doux

$$M_1 = F_1 \oslash (F_1 + F_2)$$
$$M_2 = F_2 \oslash (F_1 + F_2)$$

puis

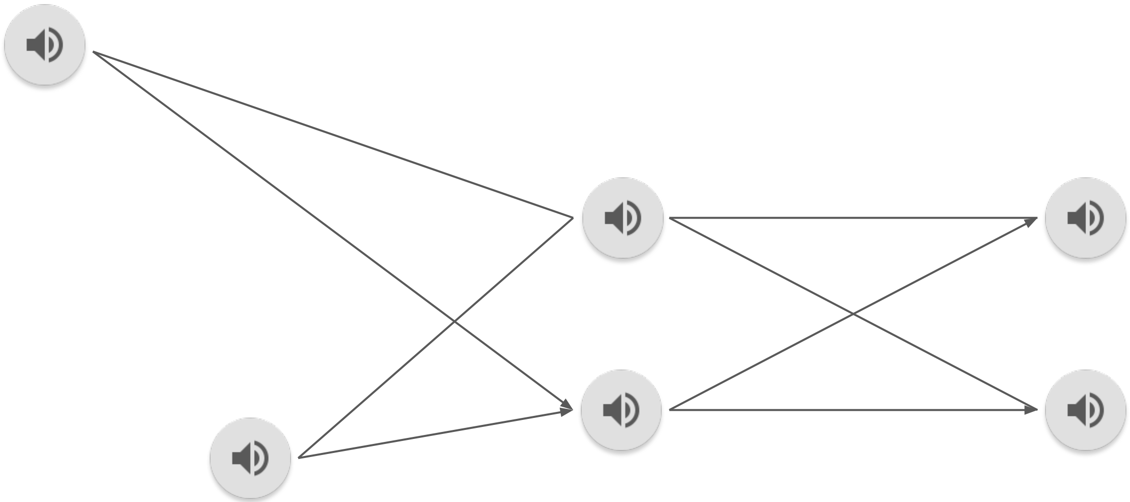
$$\hat{X}_1 = M_1 \odot Y$$
$$\hat{X}_2 = M_2 \odot Y$$

Modélisation de sources musicales

Modélisation de sources musicales

Position

ICA

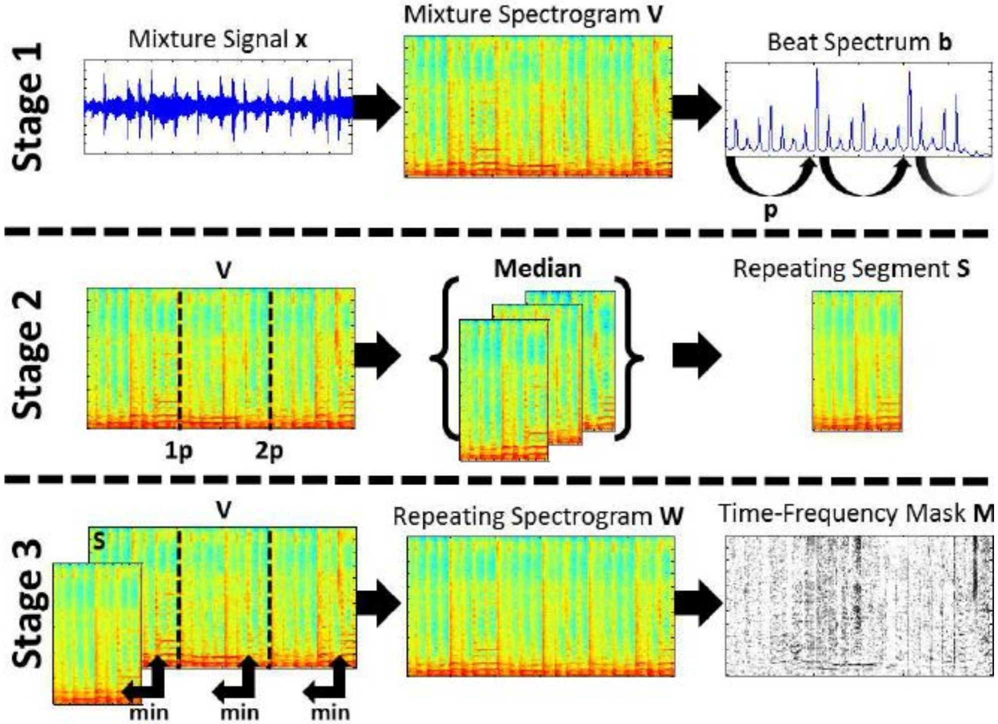


Sources *Enregistrement(s)* *Sources estimées ?*

Modélisation de sources musicales

Répétition

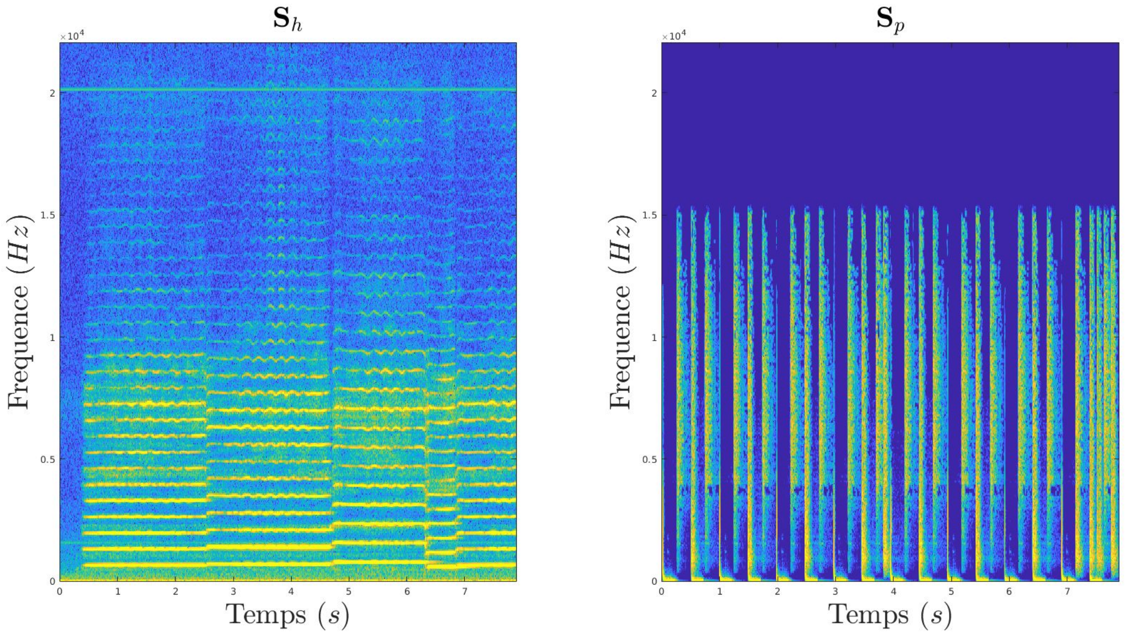
Repet



Modélisation de sources musicales

Timbre

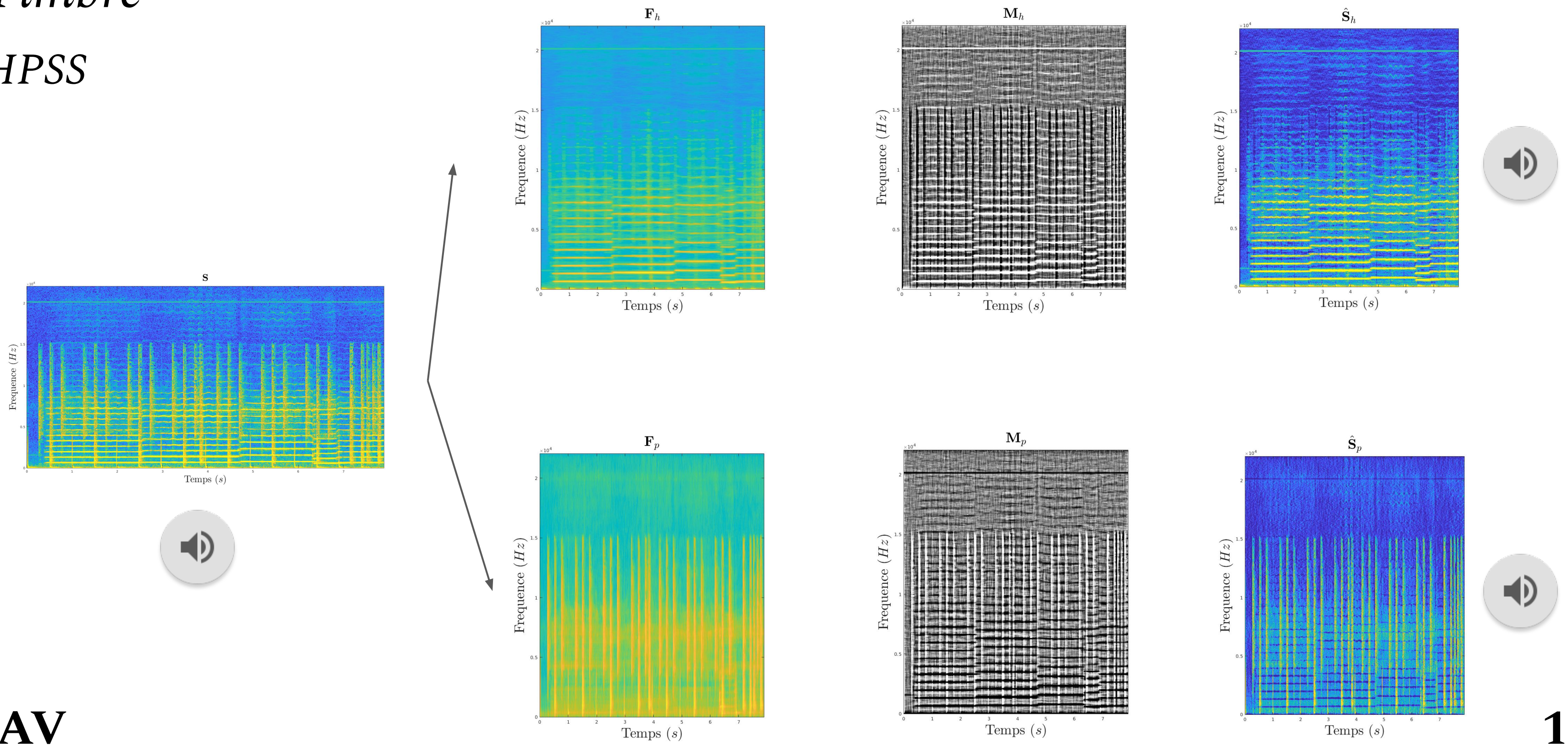
HPSS



Modélisation de sources musicales

Timbre

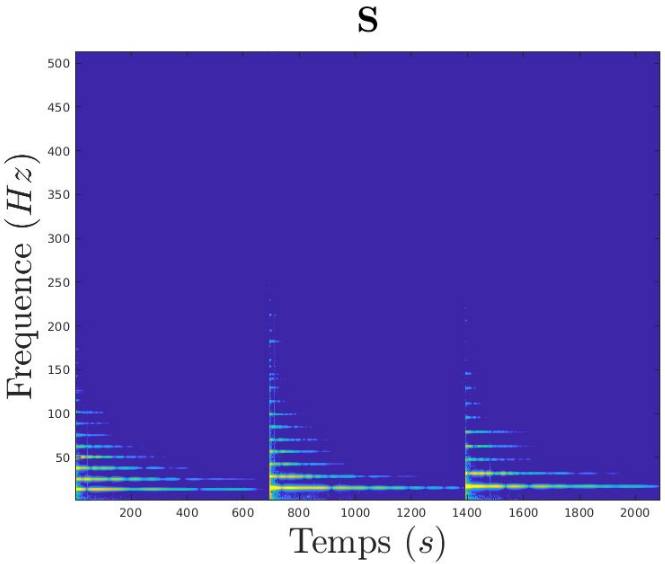
HPSS

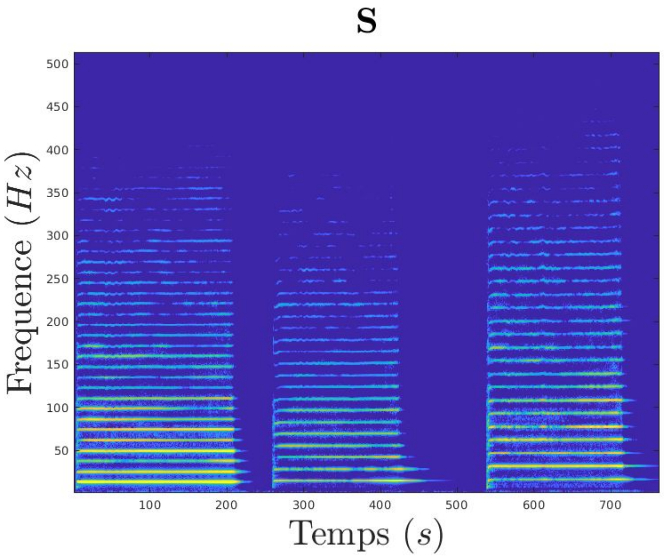


Modélisation de sources musicales

Timbre

NMF

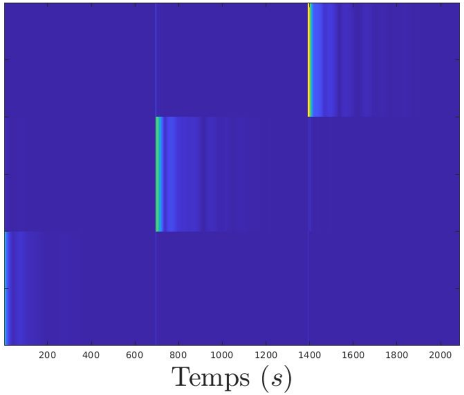


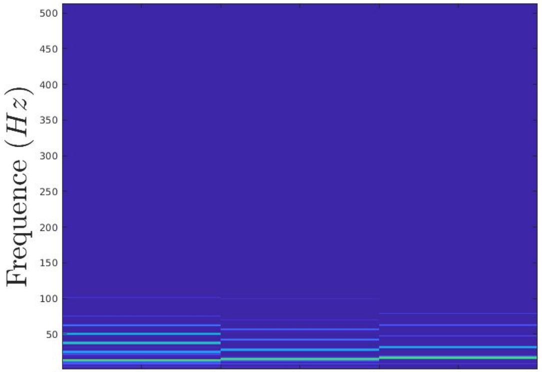


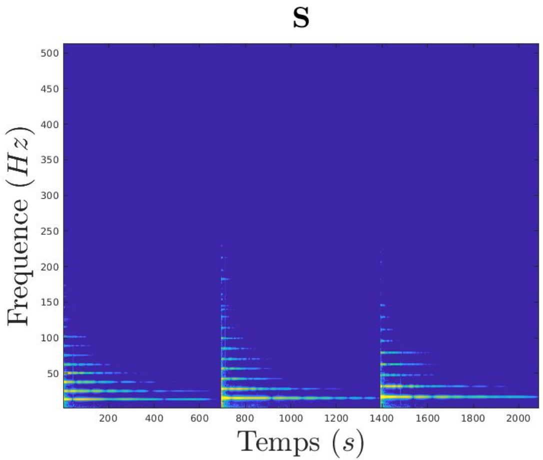
Modélisation de sources musicales

Timbre

NMF



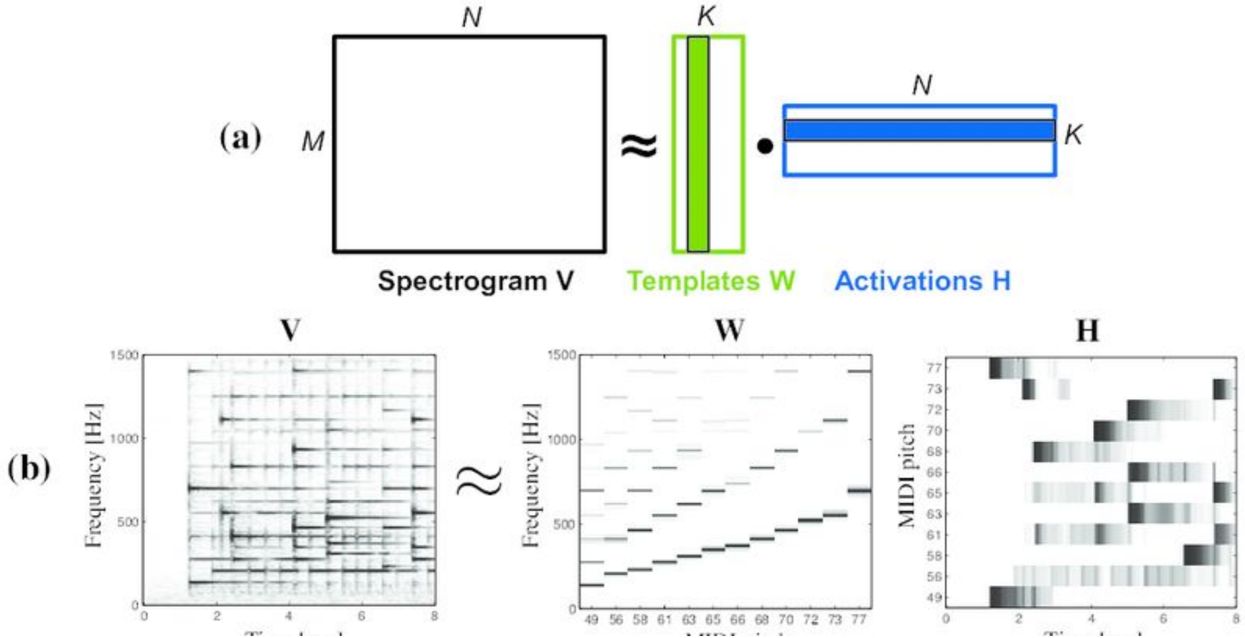




Modélisation de sources musicales

Timbre

NMF



Modélisation de sources musicales

Timbre

NMF

*On cherche **D**, **A** non négatives telles que*

$$\mathbf{DA} \approx \mathbf{S}$$

C'est un problème d'optimisation sous contraintes

$$\{\hat{\mathbf{D}}, \hat{\mathbf{A}}\} = \arg\min_{\mathbf{D},\mathbf{A}} \mathcal{E}(\mathbf{S} - \mathbf{DA}) \quad \text{s.c.} \quad \mathbf{D} \in \mathcal{M}_{K,R}(\mathbb{R}_+), \mathbf{A} \in \mathcal{M}_{R,N}(\mathbb{R}_+)$$

Modélisation de sources musicales

Timbre

NMF

Souvent, ε est une simple norme euclidienne

$$\mathcal{E}(\mathbf{S} - \mathbf{DA}) = \frac{1}{2} \|\mathbf{S} - \mathbf{DA}\|_F^2 = \frac{1}{2} \sum_i \sum_j \, [\mathbf{S} - \mathbf{DA}]_{i,j}^2$$

Malheureusement, le problème devient non convexe...

Modélisation de sources musicales

Timbre

NMF

Solution : descente de gradient alternée

$$\mathbf{D} \lhd \mathbf{D} - \eta_{\mathbf{D}} \odot \nabla_{\mathbf{D}} \mathcal{E}(\mathbf{S} - \mathbf{D}\mathbf{A})$$
$$\mathbf{A} \lhd \mathbf{A} - \eta_{\mathbf{A}} \odot \nabla_{\mathbf{A}} \mathcal{E}(\mathbf{S} - \mathbf{D}\mathbf{A})$$

Modélisation de sources musicales

Timbre

NMF

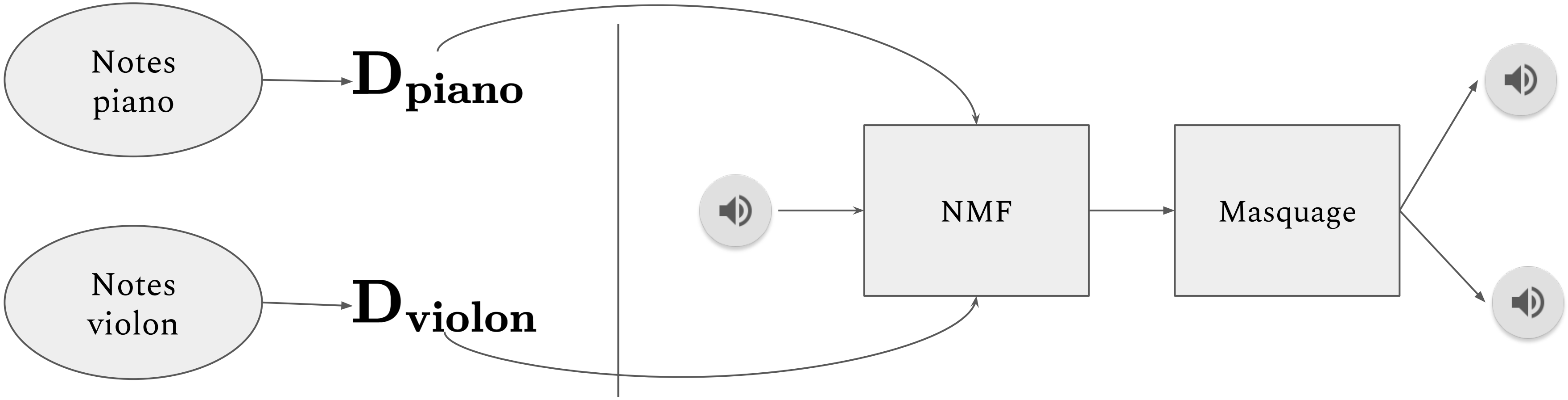
En choisissant bien les pas, on peut assurer la non-négativité de D et A. Les règles de mises à jour sont alors :

$$\mathbf{D} \triangleleft \mathbf{D} \odot \left(\mathbf{S}\,\mathbf{A}^\top\right) \oslash \left(\mathbf{D}\,\mathbf{A}\,\mathbf{A}^\top\right)$$
$$\mathbf{A} \triangleleft \mathbf{A} \odot \left(\mathbf{D}^\top \mathbf{S}\right) \oslash \left(\mathbf{D}^\top \mathbf{D}\,\mathbf{A}\right)$$

Modélisation de sources musicales

Timbre

NMF

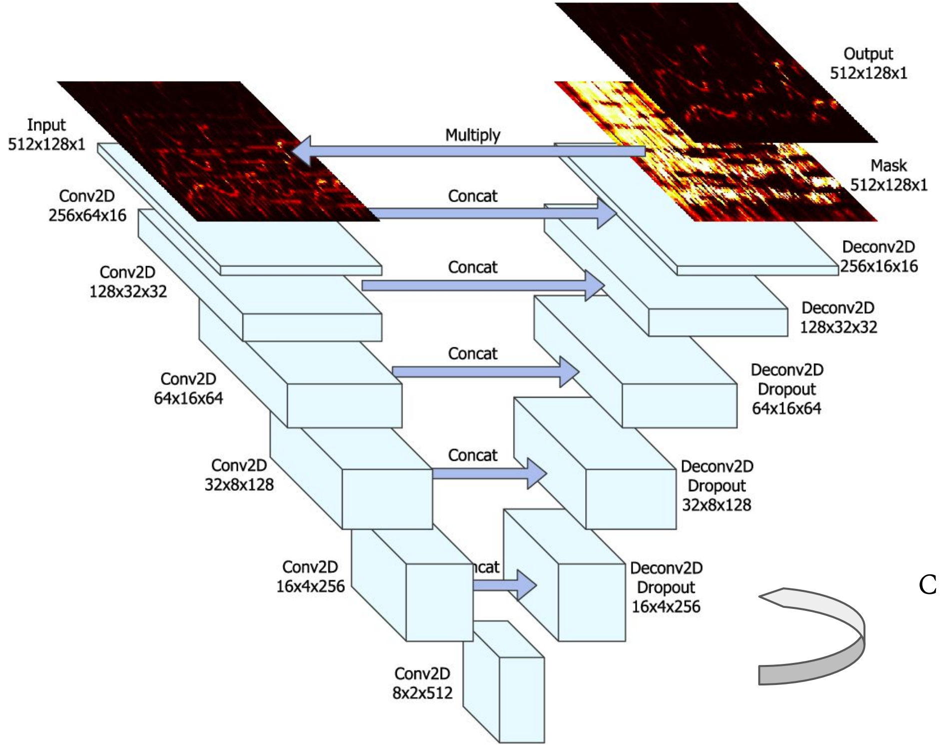


Méthodes par Deep Learning

Méthodes par Deep Learning

U-Net, le retour

Conv2D :
- Filtres : 16
- Stride : 2
- Kernel : 5



Conv2DTranspose :
- Filtres : 256
- Stride : 2
- Kernel : 5

Ressources

Ressources

- Calcul des règles de mise à jour du NMF détaillés
- Fundamentals of Music Processing Notebooks, Pr. Mueller
- Enregistrements de notes, Université d'Iowa