

MODELES SOURCE/FILTRE ET FACTORISATION EN MATRICES NON NEGATIVES



Romain HENNEQUIN

Institut Télécom, Télécom ParisTech, CNRS LTCI - 46, rue Barrault - 75634 Paris Cedex 13 - France

e-mail : hennequi@telecom-paristech.fr

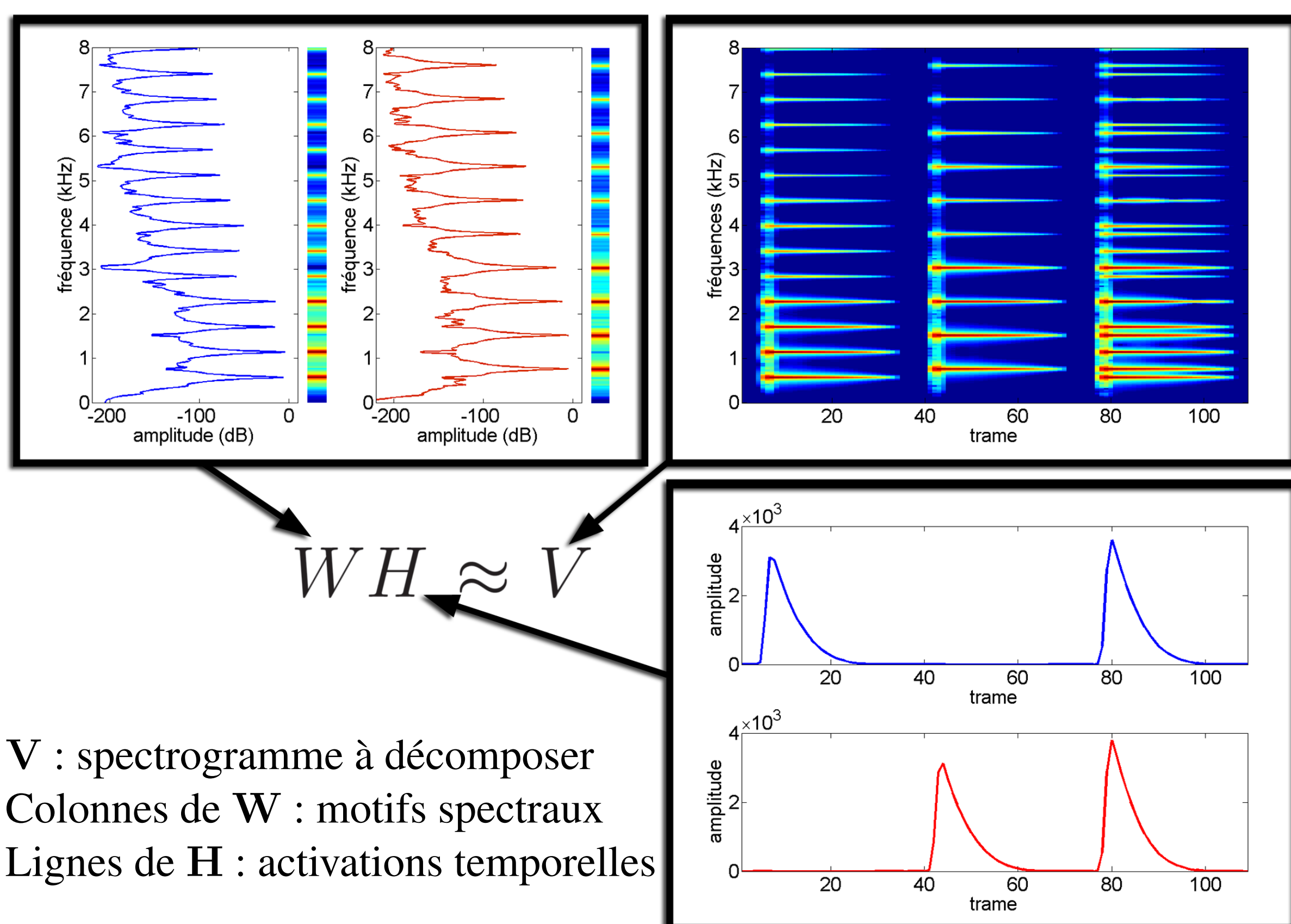
Introduction

- Contexte : décomposition de scène musicale dans une optique de séparation/transcription
- Décomposition intelligente en éléments musicaux/activités temporelles
- Introduction de modèles source/filtre dans la NMF : prise en compte des variations spectrales d'un élément musical au cours du temps.

Factorisation en matrices non-négatives (NMF)

- Puissante technique de réduction de rang de données non-négatives (Lee et Seung [1]).
- Propriété fondamentale de la NMF : contrainte de non-négativité des données
 - Combinaisons uniquement additives (pas d'énergie noire)
 - Description *perceptive* des données : décomposition des spectrogrammes musicaux sur une base de *notes*.
- Utilisation de la redondance entre les trames dans les spectrogrammes
- Application en transcription automatique [2], séparation de source [3]...
- Factorisation des spectrogrammes en motifs fréquentiels/activations temporelles :

$$\mathbf{V} \approx \mathbf{W}\mathbf{H} \quad \forall (f, t) \in \llbracket 1, F \rrbracket \times \llbracket 1, T \rrbracket \quad \mathbf{V}_{ft} \approx \sum_{r=1}^R \mathbf{W}_{fr} \mathbf{H}_{rt} \quad (FR + RT \ll FT)$$



- Approximation quantifiée par une distance (ou divergence) à minimiser entre spectre observé \mathbf{V} et reconstruit \mathbf{WH} .

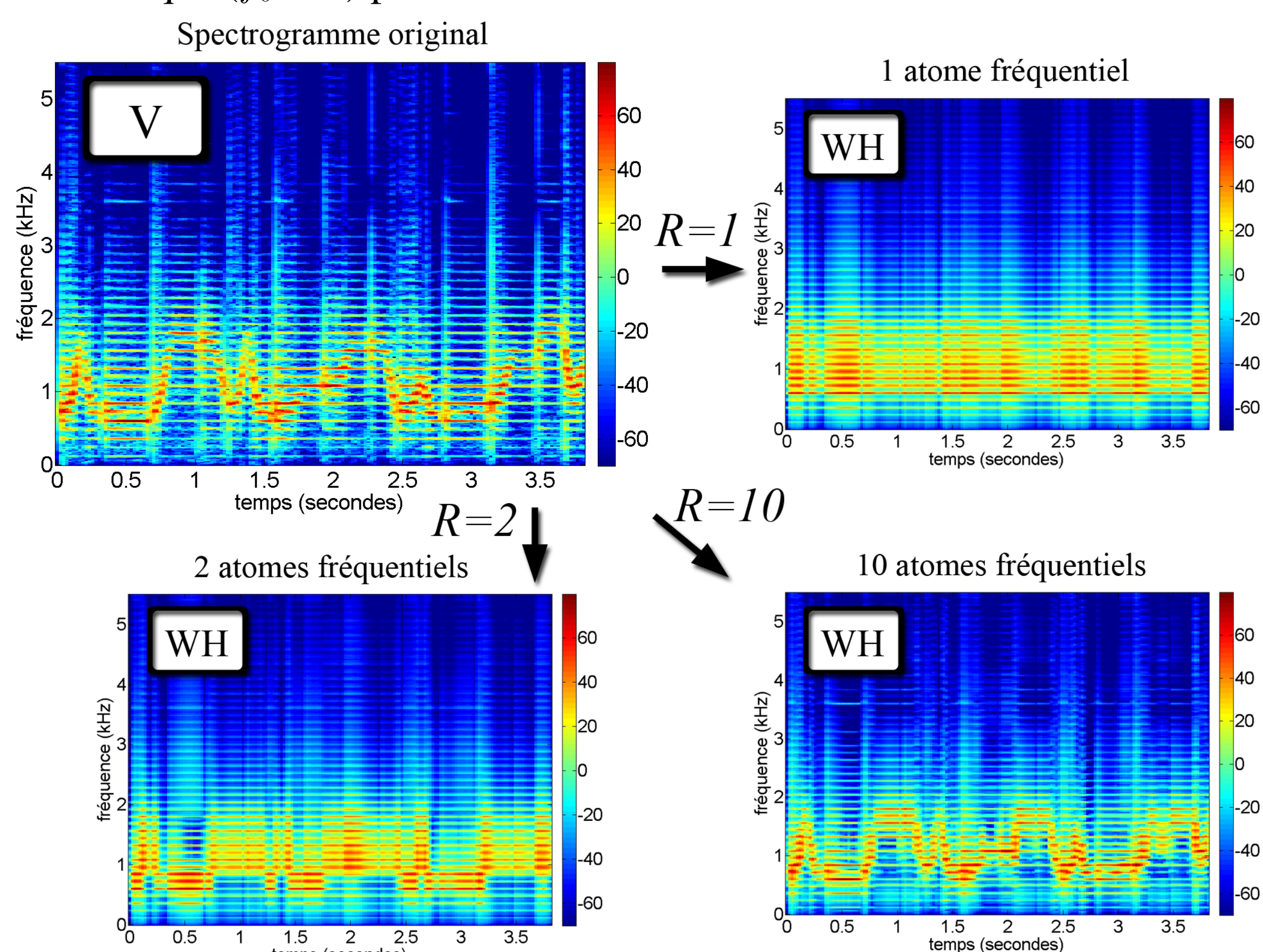
Limitation de la NMF

Fortes variations spectrales

- Pas de prise en compte de l'évolution fréquentielle de chaque note.
- Inefficace pour des sons présentant de fortes variations spectrales au cours du temps

Exemple : son de guimbarde

- Tige métallique vibrante produisant un son modulé par la bouche.
- Son harmonique (f_0 fixe) présentant une forte résonance variant au cours du temps.



Modèles source/filtre

Activation temps/fréquence

- Activations temporelles remplacées par des filtres variant dans le temps :

$$\mathbf{V}_{ft} = \sum_{r=1}^R \mathbf{W}_{fr} \mathbf{H}_{rt} \quad \Rightarrow \quad \mathbf{V}_{ft} = \sum_{r=1}^R \mathbf{W}_{fr} \mathbf{H}_{rt}(f)$$

- Limitation du nombre de paramètres : $\mathbf{H}_{rt}(f)$ doit être paramétrique.
- Interprétation avec le paradigme source/filtre : le spectre de chaque trame du signal est issu de la combinaison de motifs spectraux (sources) filtrés :
 - \mathbf{W}_{fr} : motif spectral de la source r
 - $\mathbf{H}_{rt}(f)$: filtre variant associé à la source r à l'instant t

La décomposition bénéficie de la polyvalence du modèle source/filtre qui est adapté pour de nombreux objets sonores

Paramétrisation ARMA

Forme paramétrique de $\mathbf{H}_{rt}(f)$, filtre AutoRégressif à Moyenne Ajustée (ARMA) :

$$\mathbf{H}_{rt}^{ARMA}(f) = \sigma_{rt}^2 \frac{\left| \sum_{q=0}^Q b_{rt}^q e^{-i2\pi\nu_f q} \right|^2}{\left| \sum_{p=0}^P a_{rt}^p e^{-i2\pi\nu_f p} \right|^2}$$

► $\nu_f = \frac{f-1}{2(F-1)}$: fréquence normalisée

► b_{rt}^q : coefficients de la partie MA

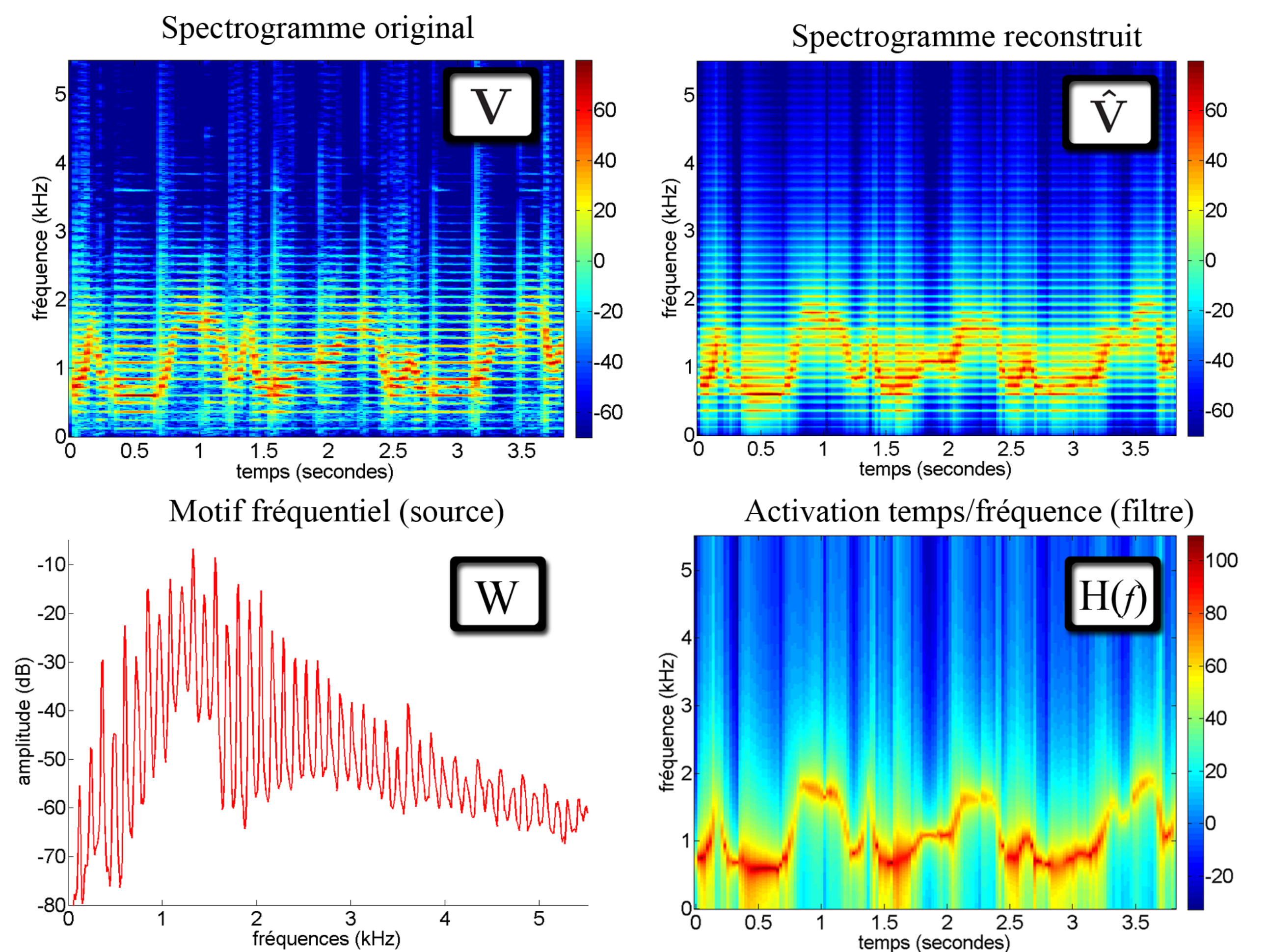
► a_{rt}^p : coefficients de la partie AR

► σ_{rt}^2 : gain global du filtre

Décomposition du spectrogramme : $\mathbf{V}_{ft} \approx \hat{\mathbf{V}}_{ft} = \sum_{r=1}^R \mathbf{W}_{fr} \mathbf{H}_{rt}^{ARMA}(f)$

Decomposition du son de guimbarde

Décomposition obtenue par un algorithme itératif (similaire à [1]).



Décomposition efficace :

- 1 seul atome fréquentiel pour un même instrument
- Résonance bien modélisée.

Conclusion

- Décomposition source-filtre : modèle de synthèse classique
- Décomposition efficace/intelligente pour des éléments à fortes variations spectrales
- Perspectives :
 - Variations de fréquence fondamentale des atomes
 - Paramétrer les variations temporelles des filtres (« continuité » trame à trame)

Bibliographie

- [1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, October 1999.
- [2] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, (New Paltz, NY), pp. 177 – 180, October 2003.
- [3] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, pp. 1066–1074, March 2007.