NMF WITH TIME-FREQUENCY ACTIVATIONS TO MODEL NON STATIONARY AUDIO EVENTS

Romain HENNEQUIN, Roland BADEAU, and Bertrand DAVID
Institut Télécom, Télécom ParisTech, CNRS LTCI - 46, rue Barrault - 75634 Paris Cedex 13 - France
e-mail: <forename>.<surname>@telecom-paristech.fr

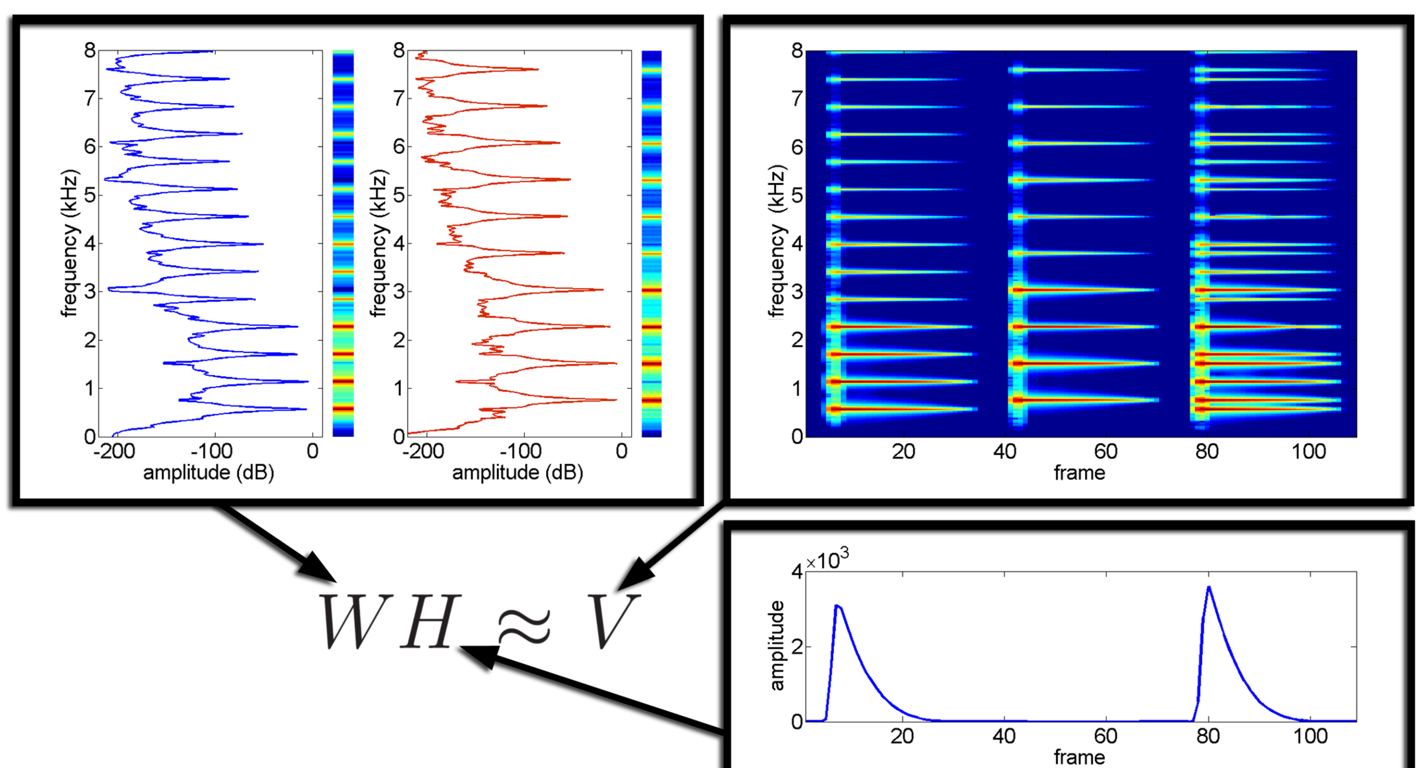
Introduction

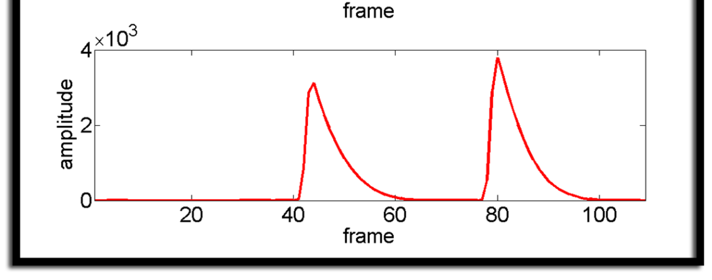
- Musical soundscape decomposition (source separation, transcription).
- Efficient decomposition into musically meaningful objects.
- Introduction of source/filter models in NMF to take highly variable spectral content within a single musical object into account.

Non-Negative Matrix Factorization (NMF)

- Powerful non-negative data rank reduction (Lee et Seung [1]).
- Fundamental property: Non-negativity constraint.
  - Only additive combination (no *black energy*).
  - *Perceptive* description: decomposition of musical spectrograms on a basis of *notes*.
- Application in automatic transcription [2], source separation [3]...
- Spectrograms factorization in frequency templates/temporal activations:

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





- V: spectrogram to be decomposed.
- Columns of W: spectral templates.
- Rows of H: temporal activations.

- Approximation quantified with a distance (or divergence) between observed spectrogram V and reconstructed spectrogram WH, to be minimized wrt W and H.

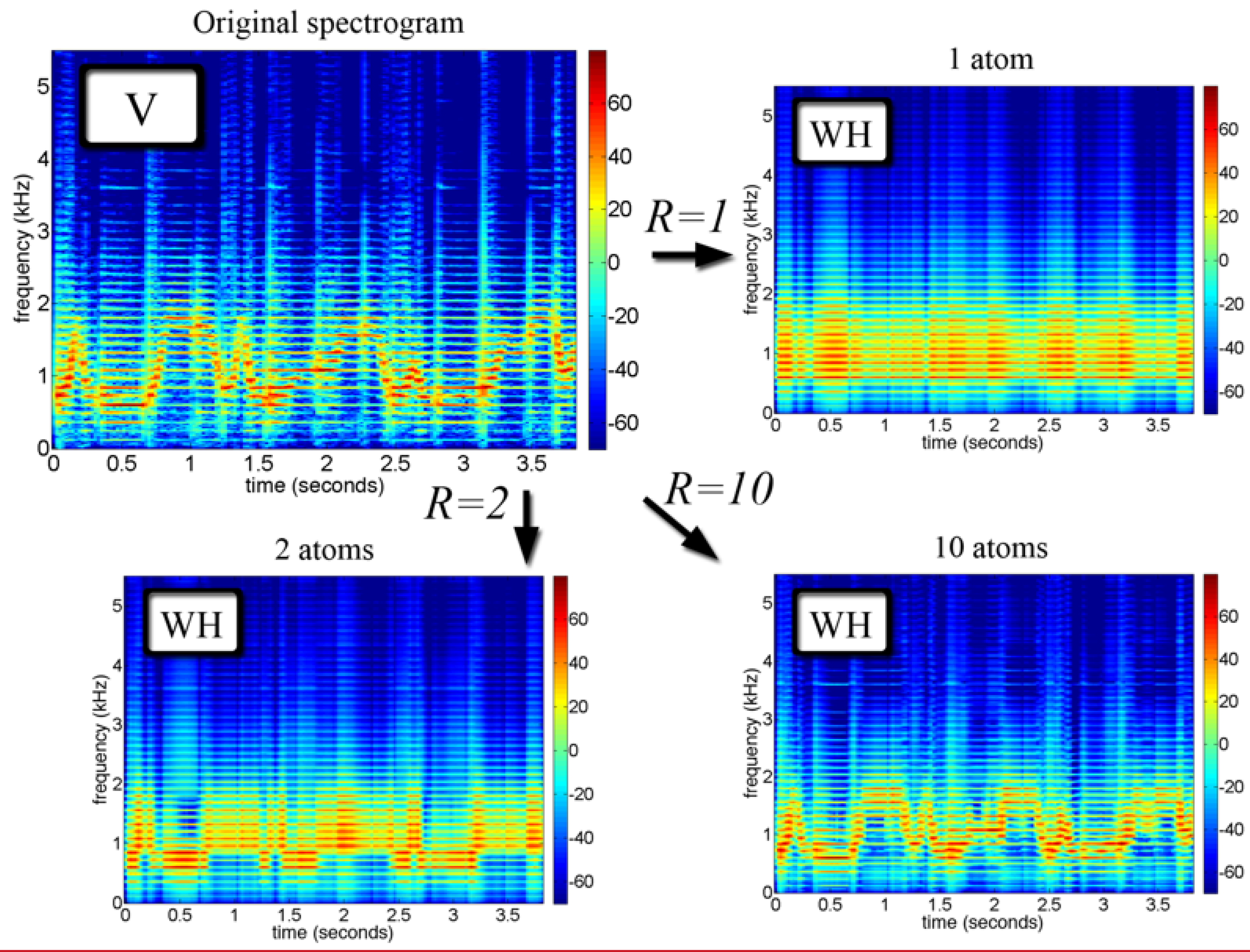
NMF issues with variable frequency content

Strong spectral variations

- Spectral variation of each note is discarded.
- Inefficient for sounds with strong spectral variations (non meaningful atoms).

Example: Jew's harp sound

- Vibrating metal tongue producing a sound modulated by the mouth of the performer.
- Harmonic sound (with fixed f_0) with a strong resonance that varies over time.



Source/filter model

Time/frequency activations

- Temporal activations replaced by time-varying filters:

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

- Limitation of the number of parameters: $\mathbf{H}_{rt}(f)$ should be parametric and smooth.
- Interpretation with the source/filter paradigm: each column of the spectrogram is a combination of filtered spectral templates:
  - $\mathbf{W}_{fr}$: spectral template of the source r.
  - $\mathbf{H}_{rt}(f)$: time-varying filter associated with the source r at time t.

The decomposition takes advantage of the versatility of the source/filter model which is well suited to model a number of different sound objects (useful for speech).

ARMA modeling

$\mathbf{H}_{rt}(f)$ chosen as an AutoRegressive Moving-Average (ARMA) filter.

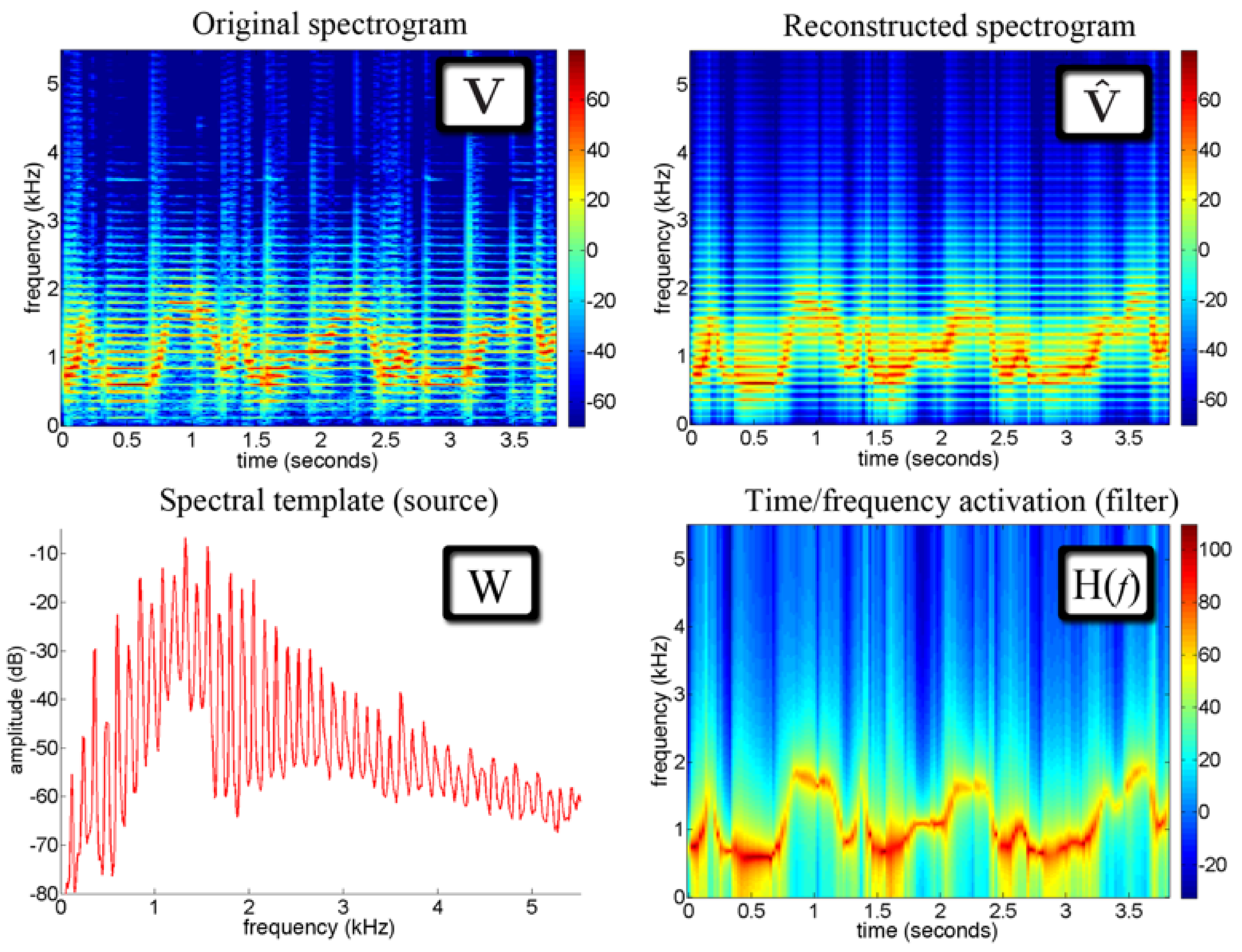
$$\mathbf{H}_{rt}^{ARMA}(f) = \sigma_{rt}^2 \frac{\left|\sum_{q=0}^{Q} b_{rt}^q e^{-i2\pi\nu_f q}\right|^2}{\left|\sum_{p=0}^{P} a_{rt}^p e^{-i2\pi\nu_f p}\right|^2}$$

- $\nu_f = \frac{f-1}{2(F-1)}$: normalized frequency.
- b_{rt}^q: coefficients of the MA part.
- a_{rt}^p: coefficients of the AR part.
- σ_{rt}^2: global gain of the filter.

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

Jew's harp sound decomposition

Decomposition obtained with a multiplicative gradient descent (similar to [1]).



Efficient decomposition:

- A single atom for a single instrument.
- Resonance accurately modeled.

Conclusion

- Source-filter model: classical synthesis model.
- Efficient decomposition for elements with strong spectral variation.
- Outlook:
  - Pitch variation of each atom.
  - Constraints on temporal variations of the filters (frame-to-frame "continuity").

Bibliography

[1] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, October 1999.

[2] P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *WASPAA*, pages 177 – 180, New Paltz, NY, October 2003.

[3] T. Virtanen. Monaural sound source separation by nonnegative matrix factorization with temporal continuity. *IEEE TASLP*, 15(3):1066–1074, March 2007.

ICASSP 2010, Dallas, Texas, USA, March 14-19, 2010