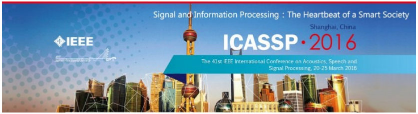


Complex NMF under phase constraints based on signal modeling

Application to audio source separation

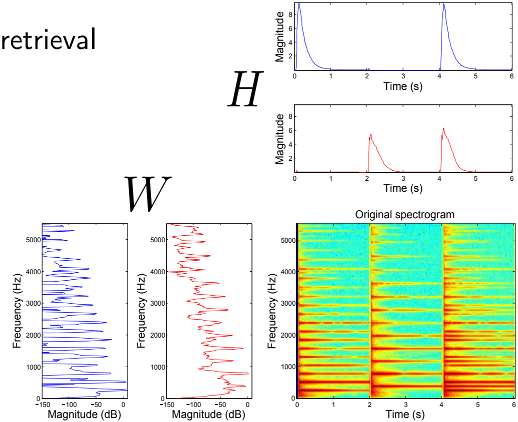
Paul Magron, Roland Badeau, Bertrand David

LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

March 23, 2016

Context

- Source separation
- NMF
- Phase retrieval



Context

- Wiener filtering commonly used
- Issues when the sources overlap in the TF domain.

How can we improve phase reconstruction in NMF-based source separation?

Wiener $\rightarrow$ CNMF-ϕ

Outline

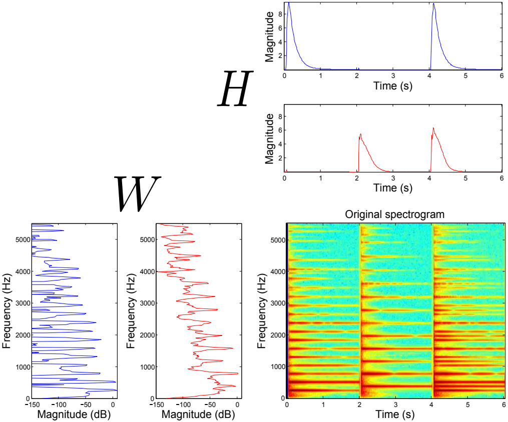
Phase reconstruction in NMF

Proposed Model

Experimental results

Phase reconstruction / NMF Model

- Non negative data: magnitude spectrogram $|X|$
- $|X| \approx WH = \sum_k W_k H_k$

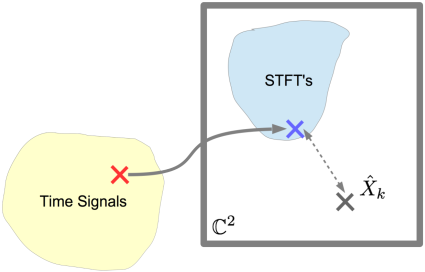


Phase reconstruction / Wiener filtering

$$\underbrace{\hat{X}_k}_{\text{Source STFT (Estimated)}} = \underbrace{\frac{W_k H_k}{\sum_{l=1}^{K} W_l H_l}}_{\text{Mask}} \odot \underbrace{X}_{\text{Mixture STFT}} \quad (1)$$

- ϕ-source = ϕ-mixture
- ⊖ Issues in sound quality when sources overlap in the TF domain
- ⊖ $\hat{X}_k \neq$ STFT of a $\hat{x}_k(n)$

Phase reconstruction / Consistency



Consistency-based approaches

- Find a $\hat{X}_k$ that is close to a STFT
- Griffin & Lim, 84 (iterative)
- Leroux, 2008, 2013

⊖ Magron, Icassp 2015, Consistency $\not\Rightarrow$ sound quality

Proposed model / Key ideas

Our approach

- Phase constraints based on **time signal properties**
- Complex NMF (CNMF) framework [Kameoka, 2009]

2 novelties

- Phase unwrapping
- Repetition of audio events

Proposed model / Background

Complex NMF (CNMF) [Kameoka, 2009]

- Mixture model:

$$\hat{X}(f,t) = \sum_{k=1}^{K} \underbrace{W(f,k)H(k,t)}_{\text{NMF model}} e^{i\phi_k(f,t)} \tag{2}$$

- Estimation by minimization of

$$\underbrace{\sum_{f,t} |X(f,t) - \hat{X}(f,t)|^2}_{\text{Distance } D(X,\hat{X})} + \sigma_s \underbrace{2\sum_{k,t} H(k,t)^p}_{\text{Sparsity penalty } C_s(H)}$$

Proposed model / Phase unwrapping

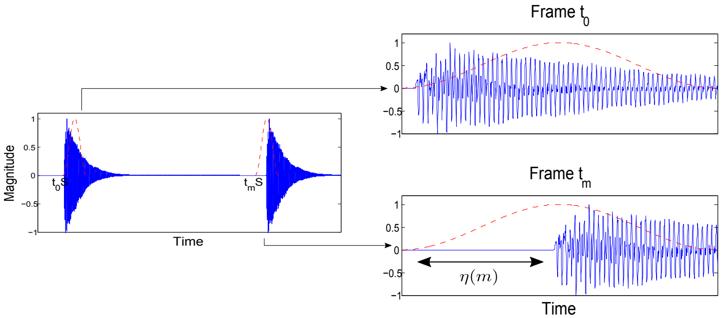
1. For each source, onset frames are detected $\rightarrow \{T_k\}$
2. Each source is modeled as a $\sum$ of sines:
   - frequency peaks are estimated with QIFFT
   - each channel f is assigned to one sine frequency $\nu_k(f)$
   - the phase in channel f is mainly governed by $\nu_k(f)$
3. phase unwrapping in channel f:

$$\Delta\phi_k(f,t) = 2\pi S \nu_k(f),$$

Unwrapping cost function:

$$\mathcal{C}_u(\phi) = \sum_{f,k} \sum_{t \neq \text{onsets}} |X(f,t)|^2 |e^{i\Delta\phi_k(f,t)} - e^{2i\pi S\nu_k(f)}|^2$$

Model of repeated audio events

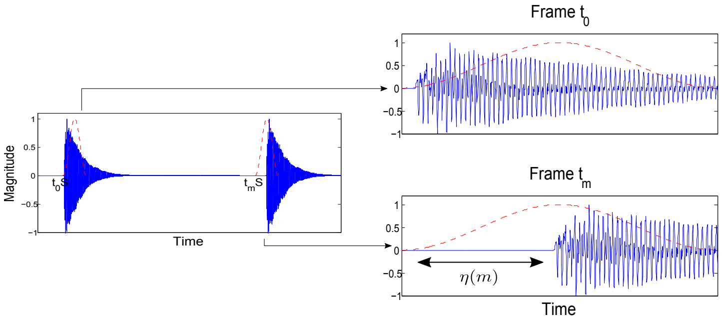


Two onset signals are equal up to a gain factor and a delay:

$$X(f, t_m) \approx X(f, t_0)\rho e^{i\lambda(m)f}, \text{ with } \lambda(m) = \frac{2\pi\eta(m)}{F}.$$

$$\underbrace{\phi(f,t)}_{\text{phase within an onset frame}} \approx \underbrace{\psi(f)}_{\text{reference phase}} + \underbrace{\lambda(t)f}_{\text{offset}}.$$

Model of repeated audio events



Repetition cost function:

$$\mathcal{C}_r(\phi, \psi, \lambda) = \sum_{f,k} \sum_{t \in \Omega_k} |X(f,t)|^2 |e^{i\phi_k(f,t)} - e^{i\psi_k(f)} e^{i\lambda_k(t) f}|^2$$

CNMF under phase constraints

Complete cost function:

$$\mathcal{C}(\theta) = \underbrace{D(X,\hat{X})}_{\text{NMF}} + \sigma_u \underbrace{\mathcal{C}_u(\phi)}_{\text{Unwrapping}} + \sigma_r \underbrace{\mathcal{C}_r(\phi,\psi,\lambda)}_{\text{Repetition}} + \sigma_s \underbrace{\mathcal{C}_s(H)}_{\text{Sparsity}}$$

- The variables are $\theta = \{W, H, \phi, \psi, \lambda\}$;
- σ_u, σ_r and σ_s are prior weights which promote the constraints.

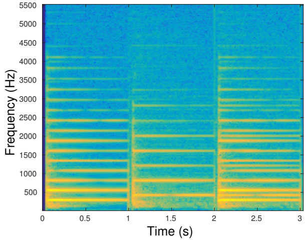
Model estimation:

Minimization of $\mathcal{C}(\theta)$.

- Coordinate descent method $\rightarrow$ Iterative procedure.
- Convergence is not guaranteed but observed in practice.

Protocol & datasets

- Synthetic mixtures of sinusoids;
- Mixtures of piano notes (MAPS database);
- $Fs = 11025$ Hz;
- The STFT uses a 46 ms-long Hann window and 75 % overlap.



Methods:

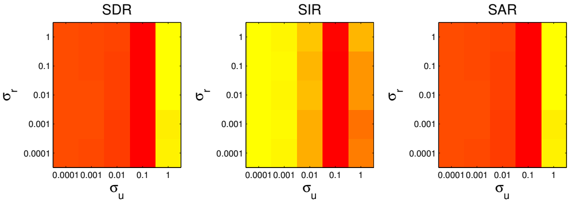
- **NMF-W**: 30 iterations of KLNMF + Wiener filtering;
- **CNMF**: 10 iterations of CNMF without phase constraints;
- **CNMF-ϕ**: 10 iterations of CNMF with phase constraints;

Score:

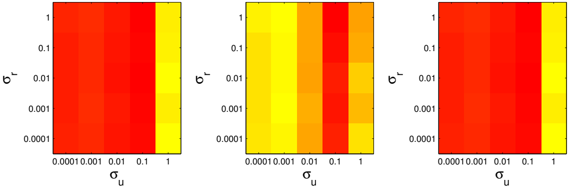
- BSS Eval [Vincent, 2006] $\rightarrow$ SDR, SIR and SAR.

Influence of the weights

- Sparsity: $p = 1$ and $\sigma_s = ||X||^2 K^{-(1-p/2)} 10^{-5}$.
- Sinusoids:

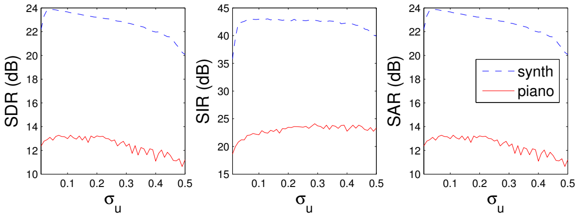


- Piano notes:



Influence of the weights

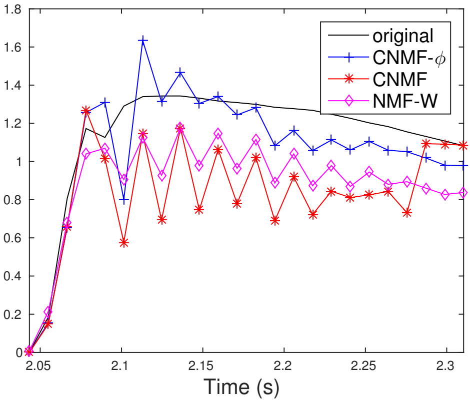
- With $\sigma_r = 0.2$:



$\rightarrow (\sigma_u, \sigma_r) = (0.2, 0.2)$ for robustness and higher scores.

Source separation

Reconstruction of a B2 piano note partial from a mixture made up of two piano notes (E2 and B2):



TELECOM ParisTech

Source separation

Separation results:

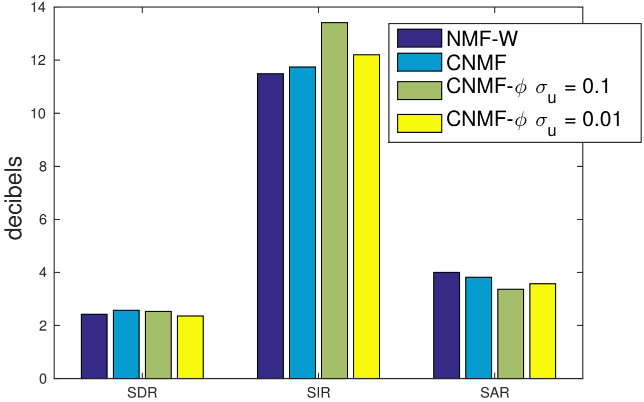
Data	Method	SDR	SIR	SAR
Synthetic sinsuoids	NMF-W	12.1	17.5	14.1
	CNMF	12.0	14.6	**16.1**
	CNMF-ϕ	**14.0**	**20.7**	15.4
Piano notes	NMF-W	12.9	23.3	14.5
	CNMF	13.5	20.0	**14.8**
	CNMF-ϕ	**14.0**	**24.0**	14.6

- Improved interference rejection.
- Slight increase of SDR.

Source separation - Realistic data

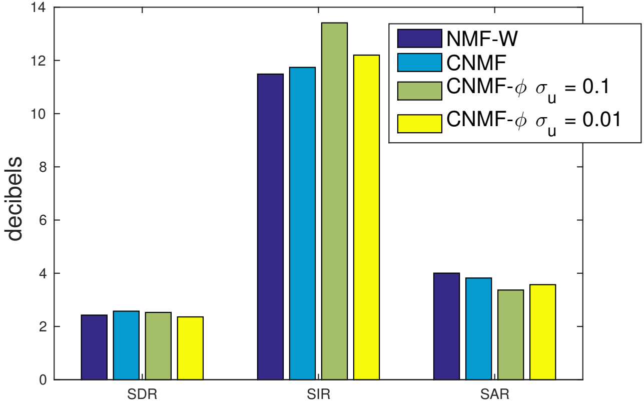
- 100 songs (rock, pop, electro...) from the Demixing Secret Database;
- The optimal weights are learned on 50 songs;
- Source separation is performed on the other 50.

Source separation - Realistic data



- Significant increase in interference rejection;
- The trade-off between SDR, SIR and SAR highly depends on the weights values.

Source separation - Realistic data



Mixture	Voice				
	Original	NMF-W	CNMF	CNMF-ϕ	CNMF-ϕ

TELECOM ParisTech

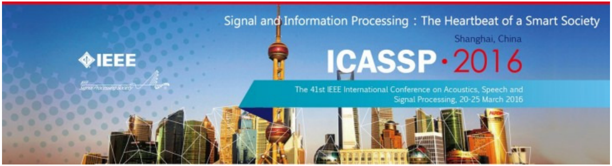
Conclusion

Complex NMF with signal model-based phase constraints

- A promising approach for separating overlapping sources in the TF domain;
- Better interference rejection than traditional Wiener filtering or unconstrained CNMF;
- The repetition constraints does not significantly improve the results.

Further work

- High sensitivity to the weight parameters;
- Optimization scheme is not efficient
  $\rightarrow$ New formulation of the problem: probabilistic framework.



Thank you!

Webpage: `http://perso.telecom-paristech.fr/~magron/`

TELECOM ParisTech