A priori probabiliste anéchoïque pour la séparation sous-déterminée de sources sonores en milieu réverbérant

Simon Leglaive, Roland Badeau, Gaël Richard
Institut Mines-Télécom - Télécom ParisTech - CNRS/LTCI - Paris, France

XXVème Colloque Gretsi, 8-11 septembre 2015, Lyon, France

Introduction

Contexte

Approche paramétrique de séparation sous-déterminée de sources sonores en milieu réverbérant.

→ A partir d'un enregistrement de musique multicanal : estimer les paramètres qui caractérisent le mélange et les sources.

Contribution

Contrainte sur l'estimation des filtres de mélange par la prise en compte d'un a priori probabiliste anéchoïque.

→ Modèle de chaîne de Markov sur la réponse en fréquence des filtres.

Méthode

Estimation des paramètres au sens du Maximum A Posteriori (MAP) par l'algorithme Espérance-Maximisation (EM).

Modèles de mélange et de sources

Mélange convolutif de J sources sur I canaux et un bruit additif, exprimé dans le domaine temps/fréquence (TF) [1] :

$$\forall (f,n) \in [\![1,F]\!] \times [\![1,N]\!], \qquad x_{i,fn} = \sum_{j=1}^{J} a_{ij,f} s_{j,fn} + b_{i,fn}$$

Modèle NMF de source

$s_{j,fn} = \sum_{k \in \mathcal{K}_j} c_{k,fn}$ — $\mathcal{K}_j$: partition de $\{1,...,K\}$ avec $K \geq J$

Composante latente Gaussienne

$c_{k,fn} \sim \mathcal{N}_c(0, w_{fk} h_{kn})$ — $w_{fk}, h_{kn} \in \mathbb{R}^+$

Filtre de mélange

$a_{ij,f}$

Bruit additif

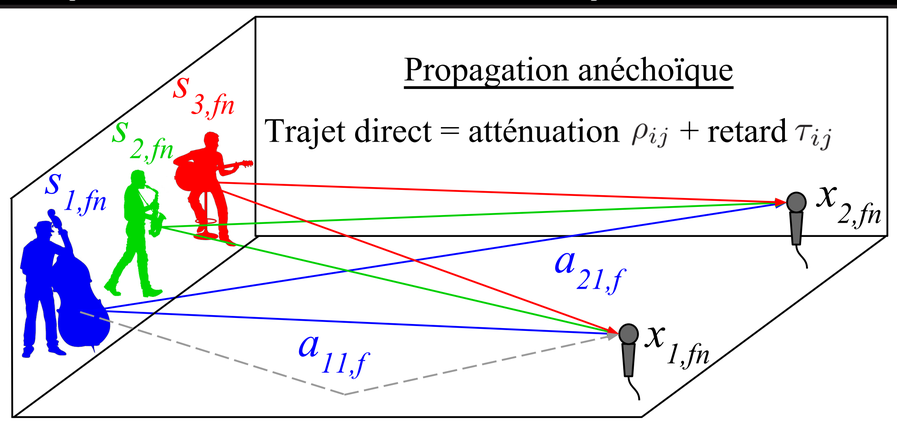
$b_{i,fn} \sim \mathcal{N}_c(0, \sigma_f^2)$

- Le problème de séparation de sources revient à estimer les paramètres

$$\boldsymbol{\theta} = \left\{ \mathbf{A} = \{a_{ij,f}\}, \mathbf{W} = \{w_{fk}\}, \mathbf{H} = \{h_{kn}\}, \boldsymbol{\Sigma_b} = \{\sigma_f^2\} \right\}.$$

- Les sources sont ensuite reconstruites par filtrage de Wiener.

A priori probabiliste anéchoïque



On néglige les réflexions sur les parois de la salle.

Modèle anéchoïque

$a_{ij,f} = \rho_{ij} \delta_{ij}^f$ où $\delta_{ij} = e^{-j2\pi\tau_{ij}} \longrightarrow a_{ij,f} = \delta_{ij} a_{ij,f-1}$

- On modélise le filtre de mélange $\{a_{ij,f}\}_{2 \leq f \leq F}$ comme un processus aléatoire suivant un modèle de chaîne de Markov :

$$a_{ij,f} = \delta_{ij} a_{ij,f-1} + b_f, \qquad \text{où } b_f \sim \mathcal{N}_c(0, \sigma_a^2).$$

- D'après ce modèle on peut écrire :

$$p(\{a_{ij,f}\}_{1 \leq f \leq F}) = p(a_{ij,1}) \prod_{f=2}^{F} p(a_{ij,f} | a_{ij,f-1}),$$

$$\text{où } p(a_{ij,f} | a_{ij,f-1}) = \frac{1}{\pi \sigma_a^2} \exp\left(\frac{-|a_{ij,f} - \delta_{ij} a_{ij,f-1}|^2}{\sigma_a^2} \right).$$

- A priori non-informatif sur $a_{ij,1}$ et $a_{ij,f}$ indépendants sur i et j :

$$-\ln p(\mathbf{A}) \stackrel{c}{=} IJ(F-1) \ln \sigma_a^2 + \frac{1}{\sigma_a^2} \sum_{f=2}^{F} \sum_{i,j} |a_{ij,f} - \delta_{ij} a_{ij,f-1}|^2.$$

Estimation des paramètres

- Maximum de Vraisemblance (MV) : sans a priori sur le mélange.
- Maximum a Posteriori (MAP) : avec a priori sur le mélange.
- Données complètes : $\left\{ \mathbf{X} = \{x_{i,fn}\}, \mathbf{C} = \{c_{k,fn}\} \right\}$.
- Algorithme EM :

→ Etape E :

$$Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{OLD}) = \mathbb{E}_{\mathbf{C}|\mathbf{X},\boldsymbol{\theta}^{OLD}} \left[-\ln p(\mathbf{X}, \mathbf{C} | \boldsymbol{\theta}) \right]$$

→ Etape M :

Estimation MV [1] :	Estimation MAP :
$\boldsymbol{\theta}_{MV}^{NEW} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ Q(\boldsymbol{\theta} \vert \boldsymbol{\theta}^{OLD})$	$\boldsymbol{\theta}_{MAP}^{NEW} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ Q(\boldsymbol{\theta} \vert \boldsymbol{\theta}^{OLD}) - \ln p(\mathbf{A})$

⇒ Les estimations MV et MAP ne diffèrent qu'au niveau de la mise à jour des réponses en fréquence $a_{ij,f}$ à l'étape M.

- Les hyper-paramètres de l'a priori δ_{ij} et σ_a^2 sont également estimés.

Expériences

Initialisation des paramètres

- $\mathbf{A}$: Algorithme de regroupement hiérarchique des points TF $\mathbf{x}_{fn}$. On en déduit une première estimation des sources par masquage TF.
- $\mathbf{W}, \mathbf{H}$: NMF avec distance de Kullback-Leibler à partir des spectrogrammes de puissance des sources précédemment estimées.
- $\boldsymbol{\Sigma}_b$: Suivant la moyenne sur les différents canaux de la variance du mélange pour chaque sous-bande fréquentielle.

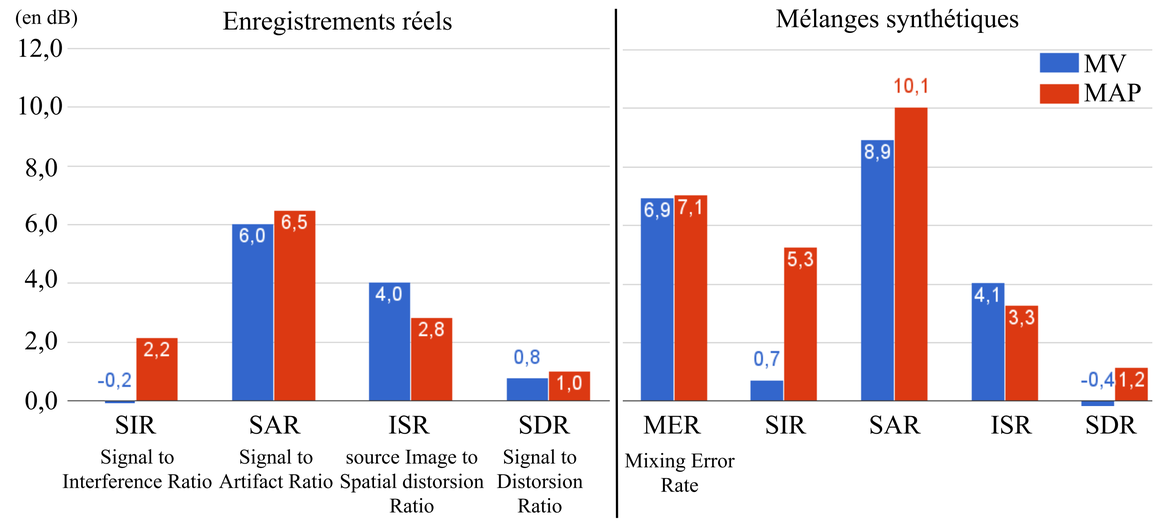
Corpus

- 5 morceaux stéréo ($\sim$ 10 s) comprenant 3 sources chacun : 2 enregistrements réels et 3 mélanges synthétiques (réponses de salle simulées).

Configuration

- $\#\mathcal{K}_j = 4$ composantes latentes pour chaque source.
- 1000 itérations de l'algorithme EM.

Résultats



Conclusion

- Introduction d'un a priori anéchoïque sous la forme d'un modèle de chaîne de Markov sur la réponse en fréquence des filtres de mélange.

- Résultats de séparation améliorés par la prise en compte de cet a priori, particulièrement au niveau du rejet des interférences.

- Pour améliorer la localisation dans le plan stéréo (augmenter l'ISR) → utiliser un modèle auto-régressif d'ordre supérieur sur les réponses en fréquence afin de capturer plus précisément le trajet direct et les premiers échos (cf. Leglaive et al., Waspaa 2015).

Référence :

[1] A. Ozerov et C. Févotte : Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation, *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3):550-563, 2010.