

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

Simon Leglaive, Roland Badeau, Gael Richard

▶ **To cite this version:**

Simon Leglaive, Roland Badeau, Gael Richard. A priori probabiliste anéchoïque pour la séparation sous-déterminée de sources sonores en milieu réverbérant. Colloque GRETSI, Sep 2015, Lyon, France. hal-01206808

HAL Id: hal-01206808
https://imt.hal.science/hal-01206808

Submitted on 29 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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, 37-39 rue Dareau, 75014 Paris, France
prenom.nom@telecom-paristech.fr

Résumé − Dans cet article, nous montrons qu'un a priori probabiliste anéchoïque sur les filtres de mélange permet d'aider la séparation aveugle et sous-déterminée de sources audio en milieu réverbérant. En considérant un modèle anéchoïque pour les filtres de mélange, la contribution de chaque source à chaque canal du mélange peut être représentée par un processus aléatoire suivant un modèle de chaîne de Markov en fréquence. Ce modèle est utilisé comme a priori pour estimer les filtres de mélange au sens du Maximum A Posteriori (MAP) en utilisant l'algorithme Espérance-Maximisation (EM). Plusieurs séparations sur des mélanges synthétiques réverbérants et sur des enregistrements réels montrent que l'estimation MAP avec a priori anéchoïque permet d'obtenir de meilleurs résultats de séparation qu'une estimation au sens du Maximum de Vraisemblance (MV) sans a priori.

Abstract − In this paper, we show that an anechoic probabilistic prior on mixing filters can help underdetermined blind source separation even in reverberant recording conditions. By considering an anechoic model for the mixing filters, we represent the contribution of each source to all mixture channels as a random process following a Markov chain model in the frequency domain. This model is used as a prior to derive a Maximum A Posteriori (MAP) estimation of the mixing filters using the Expectation-Maximization (EM) algorithm. Experimental results over reverberant synthetic mixtures and live recordings show that MAP estimation with this anechoic probabilistic prior provides better separation results than a Maximum Likelihood (ML) estimation without prior.

1 Introduction

Le problème de séparation aveugle de sources audio consiste à estimer les J signaux sources $s_j(t)$, $j = 1, ..., J$, $t = 1, ..., T$ à partir de I mélanges $x_i(t)$, $i = 1, ..., I$. Si l'on dispose de moins de mélanges que de sources ($I < J$), le problème est dit sous-déterminé. Lorsque l'on travaille sur des données audio enregistrées en milieu réverbérant, la propagation du signal source jusqu'au microphone peut être représentée par un effet de filtrage. Le mélange est alors convolutif :

$$x_i(t) = \sum_{j=1}^{J} [a_{ij} * s_j](t) + b_i(t), \tag{1}$$

où $*$ désigne le produit de convolution et $a_{ij}(t)$ est la réponse impulsionnelle de longueur L du filtre de mélange entre la source j et le microphone i. $b_i(t)$ est un bruit additif.

La plupart des approches de séparation de sources travaillent dans le domaine temps/fréquence (TF). En utilisant la Transformée de Fourier à Court-Terme (TFCT) et en supposant que la longueur du filtre de mélange L est bien plus faible que la taille de la fenêtre d'analyse utilisée, le mélange convolutif de l'équation (1) peut être approché par un mélange instantané à chaque point TF (f, n), $f = 1, ..., F$, $n = 1, ..., N$:

$$x_{i,fn} = \sum_{j=1}^{J} a_{ij,f} s_{j,fn} + b_{i,fn}, \tag{2}$$

où $x_{i,fn}$, $s_{j,fn}$ et $b_{i,fn}$, sont les TFCTs des signaux temporels associés et $a_{ij,f}$ est la transformée de Fourier discrète du filtre $a_{ij}(t)$. L'équation (2) peut être écrite de façon matricielle :

$$\mathbf{x}_{fn} = \mathbf{A}_f \mathbf{s}_{fn} + \mathbf{b}_{fn} \tag{3}$$

avec $\mathbf{x}_{fn} = [x_{1,fn}, ..., x_{I,fn}]^T$, $\mathbf{s}_{fn} = [s_{1,fn}, ..., s_{J,fn}]^T$, $\mathbf{b}_{fn} = [b_{1,fn}, ..., b_{I,fn}]^T$ et $\mathbf{A}_f = [a_{ij,f}]_{ij} \in \mathbb{C}^{I \times J}$.

Afin notamment d'améliorer les performances de séparation, un nombre croissant de méthodes guident la séparation par la prise en compte de contraintes déterministes ou d'a priori probabilistes sur les paramètres de sources et de mélange [1]. Il s'agit par exemple de supposer que la propagation entre la source j et le microphone i correspond à un retard τ_{ij} et à une atténuation ρ_{ij}. C'est le modèle anéchoïque, où les premiers échos et la réverbération sont négligés. Dans ce cas, la fonction de transfert $a_{ij,f}$ dans (2) peut être approchée par

$$d_{ij,f} = \rho_{ij} \delta_{ij}^f \quad \text{où} \quad \delta_{ij} = e^{-j2\pi\tau_{ij}}. \tag{4}$$

*Ce travail a été en partie supporté par l'Agence Nationale de la Recherche, dans le cadre du projet EDISON 3D (ANR-13-CORD-0008-02)

Dans [2], la proximité entre $a_{ij,f}$ et $d_{ij,f}$ est utilisée pour résoudre l'ambiguïté de permutation dans une approche de séparation par analyse en composantes indépendantes dans le domaine fréquentiel. Cette proximité est également utilisée dans les méthodes de séparation par masquage TF, exploitant la diversité spatiale des sources et leur parcimonie dans le plan TF [2, 3]. Dans [4], il est montré que forcer un modèle exactement anéchoïque pour le mélange ($a_{ij,f} = d_{ij,f}$) dégrade les performances par rapport au modèle convolutif (2) où aucune contrainte n'est imposée sur $a_{ij,f}$. Nous proposons ici de conserver le modèle convolutif mais de prendre en compte le modèle anéchoïque au travers d'un a priori probabiliste sur les filtres de mélange. Nous dérivons alors une estimation MAP de ces derniers, en adaptant la méthode de séparation multicanal définie dans [5] et basée sur une estimation MV des paramètres de sources et de mélange par l'algorithme EM. Nous montrons que les résultats de séparation sont alors améliorés. Dans [6], une approche similaire est utilisée pour prendre en compte des a priori probabilistes sur les matrices de covariance spatiale, dans un contexte semi-supervisé où les positions des sources et certaines caractéristiques du lieu d'enregistrement sont connues. Nous nous plaçons ici dans un scénario aveugle.

La section 2 introduit brièvement l'approche [5] sur laquelle nous nous basons. Nous montrons dans la section 3 qu'à partir du modèle anéchoïque, un filtre de mélange peut être représenté par un processus aléatoire suivant un modèle de chaîne de Markov en fréquence. Dans la section 4 nous dérivons l'estimation MAP des paramètres de mélange. Les résultats expérimentaux de séparation sont donnés dans la section 5. Enfin, nous présentons dans la section 6 la conclusion et les perspectives de ce travail.

2 Modèle et estimation MV

Nous présentons dans cette section le modèle défini par Ozerov et Févotte dans [5] et l'estimation MV des paramètres. La TFCT d'une source j est modélisée comme la somme de $\#\mathcal{K}_j$ composantes latentes Gaussiennes où $\{\mathcal{K}_j\}_{j=1}^{J}$ est une partition non-triviale de $\mathcal{K} = 1, ..., K$ avec $K \geq J$,

$$s_{j,fn} = \sum_{k \in \mathcal{K}_j} c_{k,fn} \text{ où } c_{k,fn} \sim \mathcal{N}_c(0, w_{fk}h_{kn}), \quad (5)$$

où $w_{fk}, h_{kn} \in \mathbb{R}^+$ et $\mathcal{N}_c(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ est la distribution Gaussienne complexe de densité de probabilité :

$$N_c(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\det(\pi\boldsymbol{\Sigma})} \exp[-(\mathbf{x}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})], \quad (6)$$

où $(\cdot)^H$ est le conjugué hermitien et $\det(\cdot)$ le déterminant. Les composantes $c_{k,fn}$ sont supposées mutuellement indépendantes et individuellement indépendantes sur les fréquences f et les trames n, de sorte que

$$s_{j,fn} \sim \mathcal{N}_c\Big(0, \sum_{k \in \mathcal{K}_j} w_{fk}h_{kn}\Big). \quad (7)$$

Dans (3), $\mathbf{b}_{fn}$ est un bruit blanc Gaussien, stationnaire et isotrope pour chaque bin fréquentiel f,

$$\mathbf{b}_{fn} \sim \mathcal{N}_c(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{b},f} = \sigma_f^2 \mathbf{I}_I), \quad (8)$$

où $\mathbf{I}_n$ est la matrice identité de taille n.

Soit $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{W}, \mathbf{H}, \boldsymbol{\Sigma}_\mathbf{b}\}$ avec $\mathbf{A}$ le tenseur $I \times J \times F$ d'entrées $a_{ij,f}$, $\mathbf{W}$ la matrice $F \times K$ d'entrées w_{fk}, $\mathbf{H}$ la matrice $K \times N$ d'entrées h_{kn} et $\boldsymbol{\Sigma}_\mathbf{b}$ le vecteur de taille F d'entrées σ_f^2. L'estimation MV des paramètres $\boldsymbol{\theta}$ consiste à maximiser la log-vraisemblance $\ln p(\mathbf{X}|\boldsymbol{\theta})$. Cette maximisation se fait par l'intermédiaire de l'algorithme EM dans le cas particulier des familles exponentielles. Les données complètes sont $\{\mathbf{X}, \mathbf{C}\}$, où $\mathbf{X}$ et $\mathbf{C}$ sont respectivement des tenseurs de taille $I \times F \times N$, $K \times F \times N$ et d'entrées $x_{i,fn}$, $c_{k,fn}$. L'étape E consiste à calculer l'espérance conditionnelle des statistiques suffisantes naturelles $\mathbf{R}_{\mathbf{xx},f} = \frac{1}{N}\sum_n \mathbf{x}_{fn}\mathbf{x}_{fn}^H$, $\mathbf{R}_{\mathbf{xs},f} = \frac{1}{N}\sum_n \mathbf{x}_{fn}\mathbf{s}_{fn}^H$, $\mathbf{R}_{\mathbf{ss},f} = \frac{1}{N}\sum_n \mathbf{s}_{fn}\mathbf{s}_{fn}^H$ et $u_{k,fn} = |c_{k,fn}|^2$. L'étape M consiste à réestimer les paramètres $\boldsymbol{\theta}$ en minimisant $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}')$, avec $\boldsymbol{\theta}'$ les paramètres estimés à l'itération précédente :

$$\begin{aligned} Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}') &= \mathbb{E}_{\mathbf{C}|\mathbf{X},\boldsymbol{\theta}'}[-\ln p(\mathbf{X}, \mathbf{C}|\boldsymbol{\theta})] \\ &\overset{c}{=} \sum_{f,n}\Big[\sum_k \ln(w_{fk}h_{kn}) + \sum_k \frac{\hat{u}_{k,fn}}{w_{fk}h_{kn}}\Big] \\ &+ N\sum_f \Big[\ln(\det(\boldsymbol{\Sigma}_{\mathbf{b},f})) + \mathrm{tr}\left(\boldsymbol{\Sigma}_{\mathbf{b},f}^{-1}\hat{\mathbf{R}}_{\mathbf{xx},f} - \boldsymbol{\Sigma}_{\mathbf{b},f}^{-1}\mathbf{A}_f\hat{\mathbf{R}}_{\mathbf{xs},f}^H\right. \\ &\left. - \boldsymbol{\Sigma}_{\mathbf{b},f}^{-1}\hat{\mathbf{R}}_{\mathbf{xs},f}\mathbf{A}_f^H + \boldsymbol{\Sigma}_{\mathbf{b},f}^{-1}\mathbf{A}_f\hat{\mathbf{R}}_{\mathbf{ss},f}\mathbf{A}_f^H\right)\Big], \end{aligned} \quad (9)$$

où $\overset{c}{=}$ désigne l'égalité à une constante près, $\hat{(\cdot)}$ indique l'espérance conditionnelle des statistiques suffisantes naturelles calculées à l'étape E et $\mathrm{tr}(\cdot)$ est la trace. L'algorithme EM complet est dérivé dans [5] [1]. Après estimation des paramètres, les sources sont reconstruites par filtrage de Wiener.

3 A priori probabiliste anéchoïque

Si on considère que le filtre de mélange $a_{ij,f}$ suit le modèle anéchoïque (4), on remarque que le rapport $a_{ij,f}$ sur $a_{ij,f-1}$ ne dépend pas de la fréquence et vaut δ_{ij}. Cela nous amène à modéliser le filtre de mélange $\{a_{ij,f}\}_{2 \leq f \leq F}$ comme un processus aléatoire suivant un modèle de chaîne de Markov :

$$a_{ij,f} = \delta_{ij}a_{ij,f-1} + b_f, \quad (10)$$

où b_f est un bruit blanc Gaussien complexe centré de variance σ_a^2. D'après ce modèle, nous pouvons écrire

$$p(\{a_{ij,f}\}_{1 \leq f \leq F}) = p(a_{ij,1}) \prod_{f=2}^{F} p(a_{ij,f}|a_{ij,f-1}), \quad (11)$$

$$\text{où } p(a_{ij,f}|a_{ij,f-1}) = \frac{1}{\pi\sigma_a^2} \exp\Big(\frac{-|a_{ij,f} - \delta_{ij}a_{ij,f-1}|^2}{\sigma_a^2}\Big). \quad (12)$$

1. Une version de cet algorithme avec règles multiplicatives est proposée dans [7]. Cette amélioration permet d'augmenter la vitesse de convergence.

4 Estimation MAP

L'estimation MAP des paramètres $\boldsymbol{\theta}$ consiste à maximiser la log-probabilité a posteriori $\ln p(\boldsymbol{\theta}|\mathbf{X})$. Grâce à l'algorithme EM, on obtient la même procédure qu'introduite à la section 2 et explicitée dans [5], excepté pour la mise à jour de la matrice $\mathbf{A}_f$ à l'étape M. Il s'agit en effet de minimiser par rapport à $\mathbf{A}_f$ l'espérance conditionnelle de l'opposé de la log-probabilité a posteriori des données complètes. Grâce à la règle de Bayes, ceci est équivalent à minimiser :

$$Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}') = Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}') - \ln p(\mathbf{A}), \tag{13}$$

où $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}')$ est défini à l'équation (9). En considérant un a priori non-informatif sur $a_{ij,1}$ et en supposant que les $a_{ij,f}$ sont indépendants par rapport à i et j, on calcule $-\ln p(\mathbf{A})$ grâce aux équations (11) et (12) et on obtient :

$$-\ln p(\mathbf{A}) \stackrel{c}{=} IJ(F-1)\ln\sigma_a^2 + \frac{1}{\sigma_a^2}\sum_{f=2}^{F} ||\mathbf{A}_f - \boldsymbol{\Delta}\bullet\mathbf{A}_{f-1}||_F^2, \tag{14}$$

avec $\boldsymbol{\Delta}\in\mathbb{C}^{I\times J}$ la matrice d'entrées δ_{ij}, $\bullet$ le produit matriciel terme à terme et $||\cdot||_F$ la norme de Frobenius. La nouvelle estimation de $\mathbf{A}_f$ à l'étape M est finalement obtenue en annulant le gradient de $Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}')$ par rapport à $\mathbf{A}_f$. On obtient pour $f=1$,

$$\begin{aligned}\mathrm{vec}(\mathbf{A}_f) = (\mathbf{I}_J \otimes \frac{1}{\sigma_a^2}\boldsymbol{\Sigma}_{\mathbf{b},f} + N\hat{\mathbf{R}}_{\mathbf{ss},f}^T \otimes \mathbf{I}_I)^{-1} \\ \times \mathrm{vec}\left(N\hat{\mathbf{R}}_{\mathbf{xs},f} + \frac{1}{\sigma_a^2}\boldsymbol{\Sigma}_{\mathbf{b},f}(\boldsymbol{\Delta}^*\bullet\mathbf{A}_{f+1})\right);\end{aligned} \tag{15}$$

pour $2\le f\le F-1$,

$$\begin{aligned}\mathrm{vec}(\mathbf{A}_f) = (\mathbf{I}_J \otimes \frac{2}{\sigma_a^2}\boldsymbol{\Sigma}_{\mathbf{b},f} + N\hat{\mathbf{R}}_{\mathbf{ss},f}^T \otimes \mathbf{I}_I)^{-1} \\ \times \mathrm{vec}\left(N\hat{\mathbf{R}}_{\mathbf{xs},f} + \frac{1}{\sigma_a^2}\boldsymbol{\Sigma}_{\mathbf{b},f}(\boldsymbol{\Delta}\bullet\mathbf{A}_{f-1} + \boldsymbol{\Delta}^*\bullet\mathbf{A}_{f+1})\right);\end{aligned} \tag{16}$$

pour $f=F$,

$$\begin{aligned}\mathrm{vec}(\mathbf{A}_f) = (\mathbf{I}_J \otimes \frac{1}{\sigma_a^2}\boldsymbol{\Sigma}_{\mathbf{b},f} + N\hat{\mathbf{R}}_{\mathbf{ss},f}^T \otimes \mathbf{I}_I)^{-1} \\ \times \mathrm{vec}\left(N\hat{\mathbf{R}}_{\mathbf{xs},f} + \frac{1}{\sigma_a^2}\boldsymbol{\Sigma}_{\mathbf{b},f}(\boldsymbol{\Delta}\bullet\mathbf{A}_{f-1})\right),\end{aligned} \tag{17}$$

où $\mathrm{vec}(\cdot)$ concatène les colonnes d'une matrice en un seul vecteur colonne, $\otimes$ est le produit de Kronecker et $(\cdot)^*$ l'opérateur de conjugaison.

Il nous faut également minimiser $-\ln p(\mathbf{A})$ par rapport aux paramètres δ_{ij} et σ_a^2, sachant que le modèle (4) impose $|\delta_{ij}|=1$. On trouve

$$\delta_{ij} = \frac{\sum_{f=2}^{F} a_{ij,f}a_{ij,f-1}^*}{\left|\sum_{f=2}^{F} a_{ij,f}a_{ij,f-1}^*\right|}, \tag{18}$$

$$\sigma_a^2 = \frac{1}{(F-1)IJ}\sum_{i,j}\sum_{f=2}^{F}|a_{ij,f} - \delta_{ij}a_{ij,f-1}|^2. \tag{19}$$

L'algorithme d'estimation des paramètres est ainsi très proche de celui présenté dans [5]. Seule l'étape M est modifiée au niveau de l'estimation de la matrice $\mathbf{A}_f$, afin de prendre en compte l'a priori probabiliste anéchoïque. La séparation est ici faiblement guidée, contrairement à l'approche suivie dans [6]. En effet, les hyperparamètres liés à l'a priori, δ_{ij} et σ_a^2, ne sont pas fixés mais doivent également être estimés.

5 Expériences

Dans cette section, nous comparons les performances de séparation avec et sans a priori, ce qui correspond respectivement aux estimations MAP et MV. Nous considérons des mesures de distorsion spatiale (ISR), interférences (SIR), artefacts (SAR) ainsi que le rapport signal-à-distorsion (SDR) qui est une mesure globale de performance. Lorsque les filtres de mélanges sont connus, nous calculons également le MER. Ces mesures sont définies dans [8]. Nous effectuons les expériences sur cinq morceaux de musique stéréo échantillonnés à 16 kHz, dont les caractéristiques en terme de durée, nombre de sources, type de mélange et temps de réverbération (T_{60}) sont données dans le tableau 1. Le morceau A correspond à la chanson « Sunrise » de S. Hurley utilisée dans la méthode [5]. Les morceaux $\mathrm{B}_{1,2}$ et $\mathrm{C}_{1,2}$ correspondent aux données musicales de la base de développement « dev2 » du challenge *Signal Separation Evaluation Campaign* 2008 [9] pour la tâche *under-determined speech and music mixtures*. Les mélanges synthétiques simulent une paire de microphones omnidirectionnels, séparés de 1 m et placés dans une pièce de $4,45\times3,55\times2,5$ m avec un temps de réverbération T_{60} de 130 ou 250 ms. Les enregistrements réels correspondent à la même configuration (dans le cas $\mathrm{T}_{60}=250$ ms), les signaux sources étant émis par des haut-parleurs. Pour le calcul des TFCTs, nous utilisons une fenêtre sinusoïdale de 128 ms avec un recouvrement de 50%.

L'algorithme EM est très sensible à l'initialisation des paramètres. Comme dans [4] nous choisissons d'initialiser le système de mélange $\mathbf{A}$ par l'algorithme de regroupement hiérarchique des points TF $\mathbf{x}_{fn}$ de la TFCT du mélange. Cette méthode est initialement proposée dans [10]. Grâce à cette estimation des matrices de mélange $\mathbf{A}_f$, nous en déduisons une première estimation des sources par masquage TF. Les paramètres de sources $\mathbf{W}$ et $\mathbf{H}$ sont ensuite initialisés par factorisation en matrices non-négatives des spectrogrammes de puissance des sources estimées. Nous effectuons 100 itérations des règles multiplicatives avec la divergence de Kullback-Leibler.

Nous effectuons 1000 itérations de l'algorithme EM, avec $\#\mathcal{K}_j=4$ composantes latentes pour chaque source. Notre implémentation est basée sur le code distribué avec [5] à l'adresse `http://www.irisa.fr/metiss/ozerov/`. Chaque séparation est évaluée avec et sans a priori, à partir de la même initialisation. Les résultats moyennés sur les 3

Morceaux	durée	#sources	mélange	T_{60}
A	17 s	3	synth.	130 ms
B_1 et C_1	10 s	3	synth.	250 ms
B_2 et C_2	10 s	3	réel	250 ms

TABLE 1 – Description des morceaux évalués.

	ER-MV	ER-MAP	MS-MV	MS-MAP
MER	-	-	6,94	**7,06**
SIR	-0,19	**2,16**	0,73	**5,25**
SAR	6,02	**6,50**	8,94	**10,06**
ISR	**4,04**	2,84	**4,05**	3,29
SDR	0,78	**1,01**	-0,36	**1,17**

TABLE 2 – Résultats de séparation sans (MV) et avec (MAP) a priori, moyennés sur 2 morceaux × 3 sources pour l'enregistrement réel (ER) et 3 morceaux × 3 sources pour le mélange synthétique (MS).

sources et sur l'ensemble des morceaux pour un type de mélange donné (synthétique ou réel) sont représentés dans le tableau 2. Nous observons que la prise en compte de l'a priori permet d'améliorer le MER donc l'estimation des filtres de mélange. L'augmentation du SIR pour l'estimation MAP traduit une importante réduction des interférences dans les sources estimées. Le SAR est également amélioré par la prise en comtpe de l'a priori. Les résultats concernant l'ISR montrent que l'estimation des contributions individuelles des sources au niveau de chaque canal est moins bonne. Cependant, le SDR dans le cas de l'estimation MAP étant supérieur, nous pouvons conclure que la prise en compte de l'a priori permet d'améliorer globalement la qualité de séparation.

6 Conclusion

Dans cet article, nous avons montré qu'à partir d'un modèle anéchoïque sur les filtres de mélange, il est possible de considérer que la contribution de chaque source à chaque canal suit un modèle de chaîne de Markov en fréquence. Ce modèle est pris en compte au travers d'un a priori probabiliste sur la matrice de mélange, permettant une estimation au sens du maximum a posteriori par l'algorithme EM. Nous avons montré expérimentalement que cette approche permettait d'améliorer la qualité de séparation, particulièrement au nivau du rejet des interférences.

Afin d'augmenter la mesure d'ISR et donc les performances de séparation, nous devons améliorer la localisation des sources dans le plan stéréo. Pour cela, nous pourrions envisager d'utiliser un modèle auto-régressif d'ordre supérieur sur les réponses en fréquence des filtres de mélange. Une telle approche permettrait de capturer de façon plus précise le trajet direct et les premiers échos des réponses impulsionnelles de salle.

Références

[1] E. VINCENT, N. BERTIN, R. GRIBONVAL et F. BIMBOT : From blind to guided audio source separation : How models and side information can improve the separation of sound. *IEEE Signal Processing Magazine*, 31(3):107–115, mai 2014.

[2] H. SAWADA, S. ARAKI, R. MUKAI et S. MAKINO : Grouping separated frequency components by estimating propagation model parameters in frequency-domain blind source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(5): 1592–1604, juill. 2007.

[3] O. YILMAZ et S. T. RICKARD : Blind separation of speech mixtures via time-frequency masking. *IEEE Transactions on Signal Processing*, 52(7):1830–1847, juill. 2004.

[4] N. Q. K. DUONG, E. VINCENT et R. GRIBONVAL : Underdetermined reverberant audio source separation using a fullrank spatial covariance model. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(7):1830–1840, juill. 2010.

[5] 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, mars 2010.

[6] N. Q. K. DUONG, E. VINCENT et R. GRIBONVAL : Spatial location priors for Gaussian model based reverberant audio source separation. *EURASIP Journal on Advances in Signal Processing*, 2013(1):149, sept. 2013.

[7] A. OZEROV, C. FÉVOTTE, R. BLOUET et J.-L. DURRIEU : Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation. *In Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 257–260, Prague, République Tchèque, mai 2011.

[8] E. VINCENT, S. ARAKI, F. J. THEIS, G. NOLTE, P. BOFILL, H. SAWADA, A. OZEROV, B. V. GOWREESUNKER, D. LUTTER et Q. K. DUONG, N. : The Signal Separation Evaluation Campaign (2007-2010) : Achievements and Remaining Challenges. *Signal Processing*, 92:1928–1936, mars 2012.

[9] Signal Separation Evaluation Campaign (SiSEC 2008). `http://sisec2008.wiki.irisa.fr`, 2008.

[10] Stefan WINTER, Walter KELLERMANN, Hiroshi SAWADA et Shoji MAKINO : MAP-based underdetermined blind source separation of convolutive mixtures by hierarchical clustering and ℓ1-norm minimization. *EURASIP Journal on Advances in Signal Processing*, 2007, 2007.