



Séparation aveugle de mélanges convolutifs de sources

Hakim Boumaraf

► **To cite this version:**

Hakim Boumaraf. Séparation aveugle de mélanges convolutifs de sources. MajecSTIC 2005 : Manifestation des Jeunes Chercheurs francophones dans les domaines des STIC, IRISA – IETR – LTSI, Nov 2005, Rennes, pp.316-323. inria-00000838

HAL Id: inria-00000838

<https://hal.inria.fr/inria-00000838>

Submitted on 23 Nov 2005

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.

Séparation Aveugle de Mélanges Convolutifs de Sources

Hakim Boumaraf

Laboratoire de Modélisation et Calcul
B.P. 53X, 38041 Grenoble cedex 9, France
Hakim.Boumaraf@imag.fr

Résumé : La philosophie de Séparation Aveugle de Sources consiste à estimer les signaux sources sans aucune information a priori. Dans cette contribution, nous présentons les résultats des recherches effectuées sur la séparation aveugle de mélanges convolutifs, dans le cadre de ma thèse¹ et du projet BLISS². Nous commencerons par un bref panorama sur la séparation de source. Ensuite, nous expliquerons nos algorithmes de séparation dans les cas : sans bruit et avec bruit additif. En outre, nous montrerons comment nous surmontons le problème d'ambiguïté de permutation avec deux techniques différentes. Enfin, nous donnerons les simulations des algorithmes sur des signaux de parole (non stationnaires) et des réponses impulsionnelles issues de salles acoustiques réverbérantes.

Mots-clés : séparation aveugle de sources, mélanges convolutifs, non-stationnarité, bruit additif, ambiguïté de permutation, réverbération, salles acoustiques.

1 INTRODUCTION

Depuis plus deux décennies, la Séparation Aveugle de Sources (SAS) fait l'objet de nombreux travaux en traitement de signal et de parole, en télécommunication, en biomédecine et neuroscience. Son objectif consiste à retrouver les sources originales à partir de leurs signaux mélanges sans se référer à aucune information, ni sur les sources ni sur les mélanges, hormis l'hypothèse d'indépendance statistique mutuelle des sources. En effet, le terme "aveugle" signifie qu'aucune information a priori n'est accessible. Cette problématique se présente quand plusieurs signaux (sources) se propagent dans un environnement, où ils subissent une transformation et où ils s'ajoutent pour composer les mélanges complexes, qui sont observés au niveau des capteurs de mesure. Étant donnée que la SAS modélise parfaitement beaucoup de situations physiques, celle-ci a connu des développements accrus dans plusieurs domaines d'applications. Un aperçu exhaustif sur les applications de la SAS peut être trouvé dans les articles [Mansor, 2000, Cardoso, 1998].

Historiquement, la première solution du problème de SAS a été donnée par Jutten et Herault [Herault et al., 1985], où les auteurs considèrent le cas simple dans lequel les mélanges sont instantanément établis. Leur idée est basée sur la décomposition des signaux capteurs en une somme linéaire de composantes indépendantes. Cette technique est reconnue par l'Analyse en Composantes Indépendantes (ACI). Dès lors, une variété de méthodes ont été développées pour la SAS [Hyvarinen et al., 2001, Cardoso, 1998, Abed-Meraim et al., 1997, Amari, 1998,

Mansor, 2000, Comon, 2004].

Cependant, dans beaucoup de situations pratiques, les signaux reçus aux capteurs se composent par une opération complexe, qui superpose des copies retardées des signaux sources. C'est pourquoi, dans le cas des contextes acoustiques, le modèle mathématique qui suppose que les signaux enregistrés par les microphones sont des mélanges linéaires instantanés, n'est pas valide. En effet, les mélanges "convolutifs" sont issus de la convolution des sources par les réponses impulsionnelles de l'environnement de propagation. Ce sont les caractéristiques acoustiques d'une salle qui font que les signaux microphones soient des mélanges convolutifs.

Plusieurs méthodes ont été proposées pour la séparation des sources mélangées convolutivement dans un environnement acoustique, mais une solution efficace est toujours demandée. Ceci revient à la réverbération, qui engendre plusieurs chemins de propagation et une multitude de copies retardées des signaux sources.

Nous proposons dans cet article, une contribution à la SAS par deux méthodes, qui sont opérationnelles dans les environnements acoustiques réels. L'idée de la première méthode est basée sur l'exploitation de la nonstationnarité intrinsèque des signaux. Ceci permet en fait de surmonter le problème de Gaussiennité des sources dans le domaine de Fourier [Brillinger, 1975] et d'accomplir la séparation avec les statistiques d'ordre deux. Le processus de résolution revient à effectuer une diagonalisation conjointe en utilisant un algorithme efficace [Pham, 2001], à base d'un critère d'information mutuelle. Cette idée a déjà été utilisée dans d'autres travaux [Parra and Spence, 2000, Wu and Principe, 1999, Smaragdus, 1998], mais avec des critères différents. La deuxième méthode qui prend en compte la présence du bruit dans l'environnement acoustique, quant à elle, est basée sur la maximisation de la vraisemblance. En effet, ce que nous estimons, c'est l'ensemble des statistiques qui permettent de retrouver les sources originales. L'implantation de cette démarche est réalisée par la technique Expectation-Maximization (EM). L'application première de ce travail a pour but de fournir une aide aux malentendants. La définition élémentaire de l'aide aux malentendants [Trainor et al., 2004] consiste à rendre disponible l'information sonore pour l'handicapé.

2 POSITION DU PROBLÈME

Nous considérons la situation acoustique dans laquelle M sources se propagent vers K microphones. Le modèle mathématique qui interprète le phénomène de propagation et de mélange des sources est celui qui produit les mélanges x_k , par le produit de convolution des sources s_i et les réponses impulsionnelles H_{ki} ($k=1, \dots, K$ et $i=1, \dots, M$), comme suit :

¹Remerciements sincères à M. Dinh-Tuan PHAM pour tout son apport à cette étude et pour avoir dirigé ces travaux.

²Remerciements à la consolidation du projet européen "Blind Source Separation and Applications".

$$\mathbf{x}(t)=\mathbf{H}(t)\otimes\mathbf{s}(t), \quad (1)$$

où : \otimes désigne l'opération de convolution matricielle, $\mathbf{H}(t)$ est la matrice de mélange composée des filtres H_{ki} , $\mathbf{x}(t)=[x_1(t), \dots, x_K(t)]^T$ (T désigne la transposition) est le vecteur des observations et $\mathbf{s}(t)=[s_1(t), \dots, s_M(t)]^T$ est le vecteur des signaux sources, tel que $x_k(t)=\sum_{n=-\infty}^{+\infty}\sum_{i=1}^M H_{ki}(n)s_i(t-n)$.

En disposant seulement des mélanges $\mathbf{x}(t)$, la résolution du problème de SAS consiste à retrouver un filtre de séparation $\mathbf{G}(t)$, qui permet de calculer les estimés des sources par :

$$\mathbf{y}(t)=\sum_{n=-\infty}^{+\infty}\mathbf{G}(n)\mathbf{x}(t-n), \quad (2)$$

où $\mathbf{y}(t)=[y_1(t), \dots, y_M(t)]^T$ est l'estimé du vecteur sources.

3 LES ALGORITHMES DE SÉPARATION

3.1 Algorithme pour le cas sans bruit

Comme nous l'avons expliqué dans l'introduction, le principe de base en SAS est d'arriver à concrétiser l'hypothèse d'indépendance mutuelle des sources dans le processus de séparation. En adoptant une approche de séparation au second ordre, l'idée est d'exploiter les caractéristiques intrinsèques de nos signaux afin de concrétiser l'objectif de séparation. En effet, d'une part, à partir de (2) nous pouvons voir qu'il est possible d'exprimer, à chaque fréquence, l'indépendance des sources reconstruites en fonction de leur matrice spectrale, et d'autre part, la nature non-stationnaire des signaux traités, nous permet de calculer le spectre (interspectre) des sources dans le plan temps-fréquence, en aboutissant ainsi à un ensemble de matrices, dont les termes diagonaux sont différents. C'est précisément, l'évolution du contenu fréquentiel des sources, qui va nous aider à accomplir la séparation, tout en surmontant le problème d'unicité de la solution [Belouchrani et al., 1997]. La matrice spectrale des sources estimées s'écrit comme suit :

$$S_y(t, f)=\mathbf{G}(f)S_x(t, f)\mathbf{G}^*(f), \quad (3)$$

où : * représente la transposition conjuguée d'une matrice et $\mathbf{G}(f)$ la transformée de Fourier de $\mathbf{G}(t)$.

Finalement, l'objectif revient à minimiser l'échange d'énergie entre les sources, ce qui implique que leur matrice spectrale doit s'approcher le plus possible, à chaque fréquence, d'une matrice diagonale. Afin de diagonaliser conjointement les matrices spectrales (mesure de diagonalité (4)), nous avons adopté un critère basé sur l'information mutuelle. Ce critère a été développé par Pham [Pham, 2001] :

$$\mathcal{C}(\mathbf{G}(f))=\sum_t\left\{\frac{1}{2}\log\{\det(\text{diag}[\mathbf{G}(f)\hat{S}_x(t, f)\mathbf{G}^*(f)])\}-\log\{\det(|\mathbf{G}(f)|)\}\right\} \quad (4)$$

où $\text{diag}[\cdot]$ est l'opérateur qui génère une matrice diagonale à partir des éléments diagonaux de son argument.

3.1.1 Estimation du spectre

La première étape de calcul dans notre algorithme est l'estimation de la densité spectrale de puissance (dsp) des signaux mélanges. Un large choix de ces méthodes d'estimation spectrale est possible ! (voir par exemple [Fleury, 2001]). En considérant que les signaux sont non-stationnaires et à contenu spectral variable dans le temps, notre choix est porté sur une méthode qui procède sur des blocs chevauchés. En effet, cette méthode permet d'une part, de décrire la distributivité de l'énergie en fonction du temps et de la fréquence [Max and Lacoume, 1996, pages 64-68], et d'autre part, d'estimer la dsp de l'ensemble par l'association des dsp des blocs [Max and Lacoume, 1996, page 106]. Toutefois, le choix de la taille de la fenêtre³ de la TFD (Transformée de Fourier Discrète) glissante est un paramètre essentiel pour la validité de cette approximation [Max and Lacoume, 1996, page 65].

Le périodogramme du bloc centré sur $t=\tau$, est calculé par l'équation suivante :

$$P_x(\tau, f)=\left[\sum_{t=\tau+1+N/2}^{\tau+N/2}w(t-\tau)\mathbf{x}(t)e^{-j2\pi tf/N}\right]\left[\sum_{t=\tau+1+N/2}^{\tau+N/2}w(t-\tau)\mathbf{x}(t)e^{-j2\pi tf/N}\right]^*,$$

où : N est la taille⁴ du bloc et $w(t)$ est la fenêtre d'apodisation de Tukey-Hanning, qui est utilisée pour améliorer les propriétés statistiques du spectre et réduire l'effet de bord.

Ensuite, le périodogramme est lissé par une moyenne mobile sur m points équispécés τ_1, \dots, τ_m . L'estimé du spectre est appelé : "estimateur spectral moyenné".

3.1.2 Ambiguïté de permutation

Notre approche spectrale qui présente l'avantage de ramener le problème de séparation des mélanges convolutifs à une simple diagonalisation conjointe, effectuée indépendamment, fréquence par fréquence, est malheureusement entaché d'ambiguïté.

Cela provient du fait qu'une matrice diagonale ne perd pas sa diagonalité lorsque on la multiplie à gauche par une matrice de permutation (et/ou par une matrice diagonale) et à droite par la transposée de cette matrice de permutation (et/ou diagonale). i.e. : si $\mathbf{G}(f)$ est la matrice qui diagonalise conjointement les matrices spectrales et $\mathbf{D}(f)$, $\mathbf{P}(f)$ sont respectivement une matrice diagonale et une matrice de permutation, alors : $\mathcal{C}(\mathbf{P}(f)\mathbf{D}(f)\mathbf{G}(f))=\mathcal{C}(\mathbf{G}(f))$.

Finalement, l'indétermination de la solution ne provient pas du fait de travailler fréquence par fréquence, mais revient au critère qui maximise la diagonalité des matrices spectrales. La séparation fréquence par fréquence est très importante, car l'application des procédures de séparation dans un environnement acoustique, implique que les réponses impulsionnelles des filtres de mélange soient très longues. Cela nécessite pour une approche temporelle la minimisation d'un critère contenant un très grand nombre de paramètres, à l'opposé des approches fréquentielles dont la minimisation du critère est beaucoup plus simple. L'ambiguïté de séparation de sources est un problème difficile, en particulier dans le cas d'un envi-

³pour affaiblir l'effet de bord de la troncature, il existe plusieurs formes de cette fenêtre.

⁴à choisir égale à une puissance de 2.

ronnement acoustique [Westner, 1999, Mukai et al., 2004, Mukai et al., 2001a, Mukai et al., 2001b]. Dans la littérature, nous ne trouvons que quelques idées émergées dans ce sens. Comme celle, par exemple, qui consiste à contraindre le filtre de séparation à avoir un support FIR dans le temps, car la permutation induit un filtre IIR. Celle-ci, cependant, ne convient pas aux environnements réels, qui sont caractérisés par une longue réponse impulsionnelle [Wu and Principe, 1999]. Les autres idées consistent à introduire un couplage entre les solutions dans le domaine fréquentiel [Smaragdis, 1998, Anemüller and Kollmeier, 2000], ou bien à utiliser l'hypothèse de continuité de la fonction de transfert de l'environnement de propagation [Wu and Principe, 1999, Smaragdis, 1998].

3.1.3 Estimation des profils en utilisant la SFD

Dans nos travaux [Pham et al., 2003a], nous avons commencé à résoudre le problème d'ambiguïté de la SAS par une méthode basée sur la continuité de la fonction de transfert $\mathbf{H}(f)$ (de la même façon nous imposons que $\mathbf{G}(f)$ soit continue). La puissance limitée de cette méthode dans les environnements acoustiques réverbérants, nous a amené à réfléchir sur une nouvelle idée qui soit plus performante. Nous avons donc élaboré une nouvelle procédure complémentaire à celle de la continuité du filtre de séparation (i.e. : à chaque fréquence, l'initialisation se fait par la solution précédente) [Pham et al., 2003b].

Nous allons présenter dans cette section un travail récent [Boumaraf et al., 2005], qui nous a permis d'étendre et d'améliorer l'élimination des permutations en SAS. En effet, deux nouvelles techniques ont été proposées pour améliorer l'estimation des profils d'énergie. Les deux méthodes fournissent deux façons différentes pour estimer les profils, la première réalise un lissage par moyenne locale et glissante et la deuxième emploie la Série de Fourier Discrète (SFD). Les deux techniques sont globalement comparables. La première offre un réglage fin et variable mais sensible, car il faut faire un bon compromis entre le lissage et la variance pour bien caractériser les sources par les profils. Cependant, la deuxième permet un paramétrage efficace et facile, mais pas assez fin. Seule la technique de la SFD sera présentée dans cet article.

Étant donné le type discret des signaux qu'on traite, le passage au domaine de Fourier s'effectue par la TFR (Transformée de Fourier Rapide). Donc, le spectre qui en résulte est lui aussi discret. De plus, vu la nature réelle des signaux et les propriétés de circularité et de symétrie de la TFD, une SFD est applicable sur le spectre de nos signaux. Cette SFD va permettre de trouver les signaux composites, à fréquences pures, du signal discret.

Par ailleurs, il est bien connu que les hautes fréquences d'un signal correspondent aux fortes dynamiques dans le domaine temporel. En conséquence, il est possible de réaliser un lissage par l'élimination des sinusoides à hautes fréquences, puis faire une SFD inverse. De cette manière, nous aboutissons à un lissage naturel, homogène et plus facile à contrôler.

La méthode d'élimination de permutations basée sur les profils d'énergie [Pham et al., 2003b, Boumaraf et al., 2005] consiste à admettre que l'énergie, dans le domaine des fréquences varie d'une manière proportionnelle à sa variation dans le temps. e.g. : les zones de faible énergie correspondent aux instants de silences de la parole.

Notre algorithme de SAS se résume par les étapes

suivantes :

- (i) estimation des Matrices Spectrales (MS) ;
- (ii) estimation initiale des sources par la diagonalisation conjointe des MS ;
- (iii) estimation des profils par le log de l'énergie et le lissage par SFD ;
- (iv) reconstruction des sources par la minimisation de la distance entre les sources et les profils.

3.2 Algorithme pour le cas avec bruit

Notre challenge continue, mais cette fois avec une situation plus complexe et plus difficile. Il s'agit de traiter le cas bruité de la séparation aveugle de mélanges convolutifs de sources. Dans la littérature, peu d'articles traitent ce cas. Il semblerait que cela soit dû à la difficulté d'une telle situation, en particulier pour des RSB (Rapport Signal à Bruit) élevés. e.g. : [Cardoso and Pham, 2004].

Nous présenterons dans ce qui suit la méthode développée pour traiter ce contexte difficile, qui est modélisé par l'équation suivante :

$$\mathbf{x}(t) = \mathbf{H}(t) \otimes \mathbf{s}(t) + \mathbf{b}(t), \quad (5)$$

où les vecteurs $\mathbf{x}(t)$, $\mathbf{s}(t)$, $\mathbf{b}(t)$ représentent respectivement : les mélanges, les sources et les signaux des bruits additifs aux capteurs.

Afin d'estimer les signaux sources à partir de leurs mélanges, nous avons utilisé la technique dite "EM" (pour les initiales de "Expectation-Maximization") [Dellaert, 2002], [Belouchrani, 1995, pages 44-46]. Cette approche est basée sur la minimisation itérative du maximum de vraisemblance. En séparation de sources, cette approche a été initialement employée par Belouchrani [Belouchrani, 1995, pages 44-46] dans son algorithme MLS ("Maximum Likelihood Separation"), mais ce développement ne s'appliquait qu'à la séparation des mélanges instantanés [Belouchrani and Cardoso, 1994, Belouchrani and Cardoso, 1995]. Récemment, cette méthode a été reprise dans le travail de Cardoso et Pham [Cardoso and Pham, 2004], où elle a été réétudiée et optimisée pour l'ICA. Par ailleurs, dans la littérature, cette technique apparaît aussi dans d'autres travaux, e.g. : [Welling and Weber, 2001, Ghahramani and Hinton, 1997].

Dans le cadre de ma thèse, nous avons proposé une extension de l'algorithme EM au cas des mélanges convolutifs bruités.

3.2.1 Principe de l'approche

En partant de notre modèle (5), une transposition dans le domaine de Fourier va transformer le produit de convolution en un produit normal. Ainsi, nous obtenons un modèle plus simple qui s'écrit comme suit :

$$\mathbf{x}(f) = \mathbf{H}(f) \mathbf{s}(f) + \mathbf{b}(f), \quad (6)$$

où : $\mathbf{x}(f)$, $\mathbf{s}(f)$ et $\mathbf{b}(f)$ sont respectivement les transformées de Fourier de $\mathbf{x}(t)$, $\mathbf{s}(t)$ et $\mathbf{b}(t)$.

À partir de l'équation (6) nous pouvons calculer la matrice spectrale (ou par équivalence : la matrice de covariance) du vecteur des mélanges :

$$\mathbf{R}_x = S_x(t_i, f) = \mathbf{H}(f) S_s(t_i, f) \mathbf{H}^*(f) + S_b(t_i, f) \quad (7)$$

Ayant cet ensemble de données, le problème revient à estimer les sources à partir des paramètres de la formulation de notre modèle (6). Nous nous retrouvons donc avec le problème statistique d'estimation des paramètres suivant⁵ : la matrice de mélange \mathbf{H} , les matrices de covariance des sources $\mathbf{D}_1, \dots, \mathbf{D}_r$ (r représente le nombre des blocs) et les matrices de covariances du bruit $\mathbf{N}_1, \dots, \mathbf{N}_r$. Par ailleurs, nous savons que la fonction log-vraisemblance d'un ensemble de matrices de covariance $\Gamma_1, \dots, \Gamma_r$ ($\Gamma_i = \mathbf{H}\mathbf{D}_i\mathbf{H}^* + \mathbf{N}_i$ est issue du modèle (6)), pour lesquelles nous disposons des estimés $\mathbf{R}_1, \dots, \mathbf{R}_r$, est :

$$C - \frac{1}{2} \sum_{i=1}^r w_i \{ \log \det \Gamma_i + \text{tr}(\Gamma_i^{-1} \mathbf{R}_i) \},$$

où : C est une constante et w_i est un facteur proportionnel à la taille des données (blocs). tr désigne la trace d'une matrice.

La méthode du maximum de vraisemblance est ainsi utilisée pour estimer l'ensemble \mathbf{H} , $\mathbf{D}_1, \dots, \mathbf{D}_r$ et $\mathbf{N}_1, \dots, \mathbf{N}_r$, en minimisant le critère suivant :

$$\sum_{i=1}^r w_i \{ \log \det(\mathbf{H}\mathbf{D}_i\mathbf{H}^* + \mathbf{N}_i) + \text{tr}[(\mathbf{H}\mathbf{D}_i\mathbf{H}^* + \mathbf{N}_i)^{-1} \mathbf{R}_i] \}$$

3.2.2 L'algorithme EM

Expectation-Maximization est une méthode itérative d'estimation par maximisation de la vraisemblance. Son principe de base est l'extension de l'ensemble des données observées ("*données incomplètes*") avec d'autres variables additionnelles ("*données manquantes ou complémentaires*"), pour accomplir l'estimation simplement en deux étapes itératives. La théorie a montré que l'algorithme EM a la propriété d'avoir une bonne monotonie, qui assure la croissance itérative de la vraisemblance jusqu'à un point critique [Belouchrani, 1995, pages 44-46]. Parmi les premiers articles qu'on trouve dans la littérature sur le développement de l'algorithme EM, le papier de Hartley [Hartley, 1958]. Cependant, la formulation de cet algorithme ainsi que la démonstration de sa convergence sont apparues dans l'article de Dempster [Dempster et al., 1977].

Dans l'objectif d'estimer les paramètres \mathbf{H} , $\mathbf{D}_1, \dots, \mathbf{D}_r$ et $\mathbf{N}_1, \dots, \mathbf{N}_r$ en utilisant la méthode Expectation-Maximization et ainsi solutionner le problème de séparation des mélanges convolutifs bruités, nous élargissons l'ensemble de nos données observées (incomplètes) par d'autres données nécessaires (complémentaires) au procédé de notre algorithme. À cette fin, nous considérons le modèle (complet⁶), où les vecteurs \mathbf{s}_i sont supposés observables. En conséquence, nous aurons accès à l'ensemble des matrices de covariance \mathbf{R}_i^s de \mathbf{s}_i et d'intercovariance $\mathbf{R}_i^{x,s}$. La fonction log-vraisemblance du modèle complet est ainsi proportionnelle à :

$$-\sum_{i=1}^r \{ \log \det \mathbf{D}_i + \text{tr}(\mathbf{D}_i^{-1} \mathbf{R}_i^s) + \log \det \mathbf{N}_i + \text{tr}[\mathbf{N}_i^{-1} (\mathbf{R}_i - 2\mathbf{R}_i^{x,s} \mathbf{H}^* + \mathbf{H} \mathbf{R}_i^s \mathbf{H}^*)] \} \quad (8)$$

En résumé, on obtient l'algorithme suivant :

⁵pour simplifier la notation nous omettons la variable f et le temps (blocs) est indiqué par l'indice i .

⁶on dit modèle complet car il permet d'associer (compléter) les données manquantes aux données observées (incomplètes).

- (i) Initialisation ;
- (ii) étape 'E' : estimation des statistiques complémentaires $\hat{\mathbf{R}}_i^{x,s}$, $\hat{\mathbf{R}}_i^s$;
- (iii) étape 'M' : mise à jours des nouvelles valeurs $\hat{\mathbf{D}}_i^{\text{new}}$, $\hat{\mathbf{H}}^{\text{new}}$ et $\hat{\mathbf{N}}^{\text{new}}$;
- (iv) aller à (v) si convergence, sinon retourner à (ii) ;
- (v) estimation des sources séparées.

4 SIMULATIONS NUMÉRIQUES

Afin de valider nos algorithmes, nous avons choisi d'expérimenter avec un jeu de données de réponses impulsionnelles réelles. Ces données ont été mesurées (pour le projet BLISS) par l'Université McMaster dans des salles acoustiques réverbérantes [Trainor et al., 2004]. Nous présentons dans ce qui suit deux simulations, une dans le cas avec bruit et l'autre dans le cas sans bruit. Les réponses impulsionnelles utilisées ont été mesurées dans une salle qui fait $3,4\text{m} \times 3,4\text{m} \times 2,6\text{m}$.

4.1 Mesure de la performance

La mesure de performance est une tâche essentielle en séparation de sources, plusieurs définitions de cette mesure sont données dans la littérature [Amari and Cichocki, 1998, Ziehe and Müller, 1998, Kisilev et al., 2003, Mukai et al., 2004]. Pour évaluer la performance de séparation de nos algorithmes, nous avons adopté des mesures appropriées pour chaque cas. Dans le cas sans bruit, la mesure consiste à comparer l'ordre de grandeur de la diagonale du filtre global ($\mathbf{G}(f)\mathbf{H}(f)$) à celui de son anti-diagonale. La façon naturelle d'effectuer cette comparaison est de calculer le rapport entre le produit des éléments diagonaux et le produit des éléments anti-diagonaux. Nous pouvons donc définir cet indicateur (pour le cas de 2 sources, 2 capteurs) comme suit :

$$r(f) = \sqrt{\frac{|(\mathbf{GH})_{12}(f)(\mathbf{GH})_{21}(f)|}{|(\mathbf{GH})_{11}(f)(\mathbf{GH})_{22}(f)|}} \quad (9)$$

où $(\mathbf{GH})_{ij}$ est l'élément ij du filtre global.

Cet indice indique l'occurrence des permutations quand il devient infiniment grand⁷. Pour une indication claire, nous traçons aussi sur le même graphique l'inverse de $r(f)$. Ainsi, les fréquences auxquelles la séparation est de moins bonne qualité, sont identifiées par une valeur d'indice plus loin de 0, et une permutation est identifiée par un basculement opposé entre l'indice et son inverse.

Par ailleurs, notre indice possède l'avantage d'être inchangé par rapport à l'ambiguïté d'échelle, d'une fréquence à une autre.

Dans le cas bruité, l'évaluation de la performance de séparation est effectuée par l'indice qui estime le rapport signal à bruit dans le plan temps-fréquence. Ceci permet de montrer l'affectation du contenu fréquentiel de la source séparée par le bruit. La formulation mathématique de cet indice (pour une source s_j) est naturellement donnée par :

$$\text{RSB}_j(t, f) = 10 \log \frac{S_{s_j}(t, f)}{S_{[\hat{s}_j - s_j]}(t, f)} \quad (10)$$

⁷pour la bonne représentation de l'indice, toutes les valeurs infiniment grandes sont ramenées à 1 (nous utilisons $\min(r, 1)$).

4.2 Simulation dans le cas avec bruit

Dans cet expérience, l'algorithme de SAS est appliqué sur l'ensemble de mélanges issus de la propagation d'une source à 4 capteurs. La source à séparer est un signal de parole d'environ 2 secondes. La fréquence d'échantillonnage des signaux (sources et réponses impulsionnelles) est égale à 11,025kHz. Les réponses impulsionnelles des filtres de mélange sont tronquées à 512 retards.

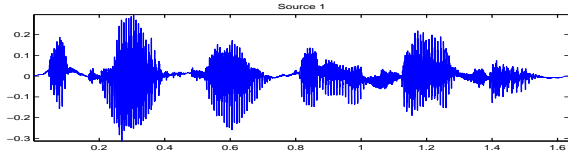


FIG. 1 – Source originale.

L'estimation des matrices spectrales est faite en procédant par des blocs de taille 2048, inter-chevauchés à 75% et moyennés sur 5 blocs. Le vecteur des mélanges est représenté sur la figure 2. Le bruit considéré est Gaussien et les taux RSB valent respectivement : -1db, -2db, -3db, -4db.

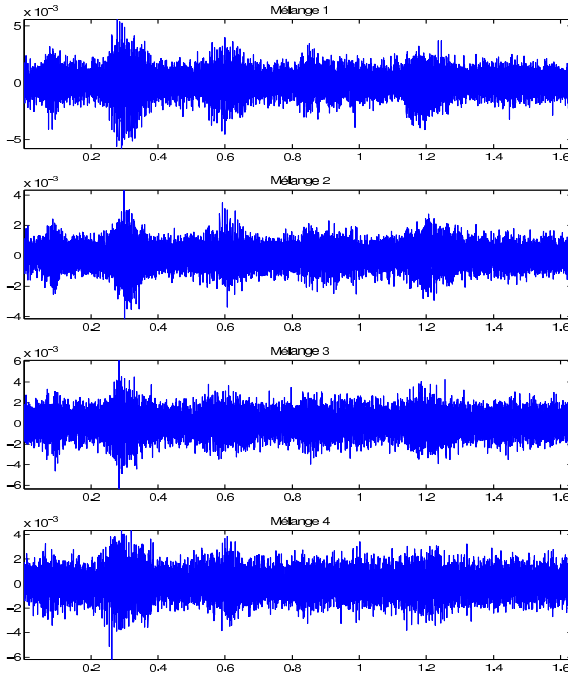


FIG. 2 – Mélanges.

Sur la figure 3 nous présentons la source séparée par l'algorithme EM. Nous pouvons voir que le bruit est éliminé et l'allure ("les bouffées") de la source originale est bien retracée par le processus de séparation.

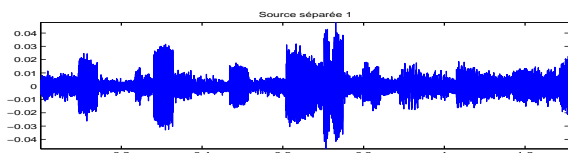


FIG. 3 – Source séparée.

Les filtres de mélange de l'environnement acoustique réel, dans lequel la source s'est propagée jusqu'aux capteurs, sont tracés sur la figure 4. La réverbération acoustique se

présente sur leurs réponses impulsionnelles par des pics d'échos.

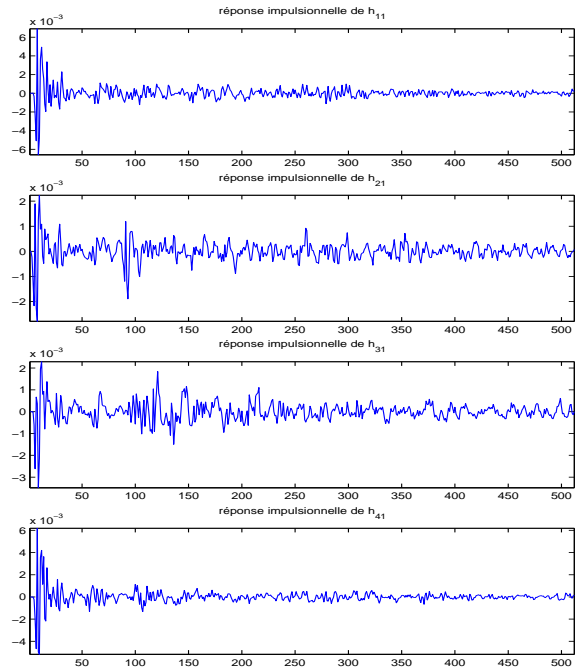


FIG. 4 – Filtre de mélange.

Le rapport signal à bruit évaluant la séparation est donné par la figure 5. Il révèle la faible affectation de la source séparée par le bruit sur tout l'horizon temps-fréquence.

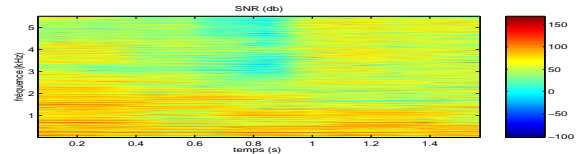


FIG. 5 – Rapport signal à bruit.

4.3 Simulation dans le cas sans bruit

Quand à cet expérience, les deux sources à séparer ont une durée d'environ 3 secondes. Les réponses impulsionnelles utilisées dans cet expérience (Figure 6) ont 512 retards. Tous les signaux sont échantillonnés à 11,025kHz.

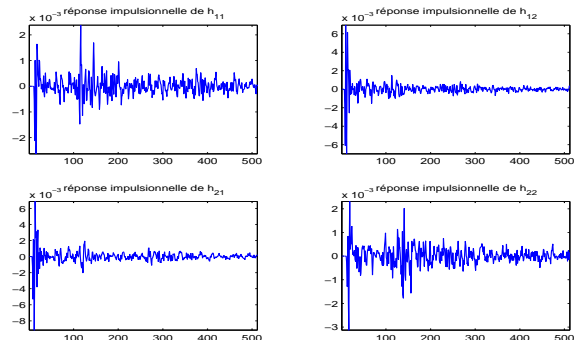


FIG. 6 – Réponses impulsionnelles du filtre considéré.

Pour l'estimation spectrale, nous avons pris comme taille de blocs 2048, comme taux de chevauchement 75% et comme points de lissage 5. Cela nous a fait 57 blocs.

Sur la figure 7 est tracé $\min(r, 1)$ et $\min(1/r, 1)$ en fonction de la fréquence. Nous voyons clairement l'élimination des permutations initiales.

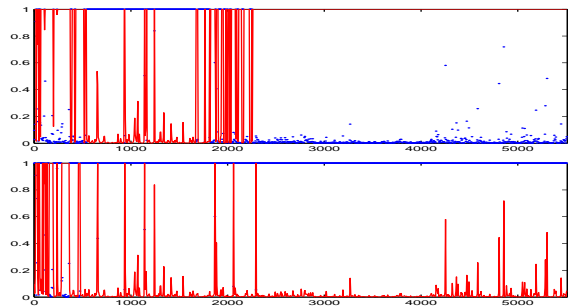


FIG. 7 – Indices de séparation (en trait continu) et son inverse (en points) avant (en haut) et après (en bas) l'élimination des permutations par la technique de la SFD.

Sur la figure 8 sont tracés les signaux des sources originales, des mélanges et des sources séparées.

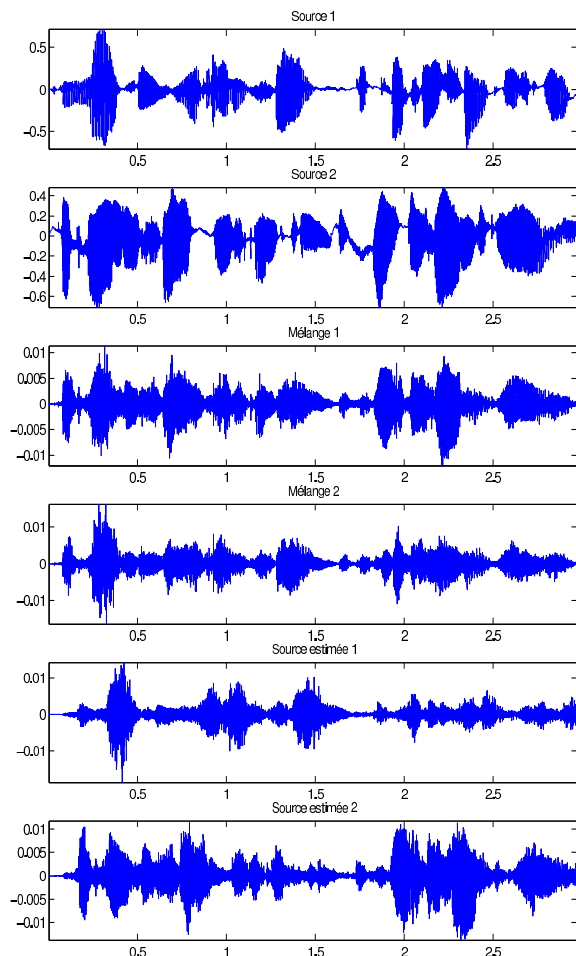


FIG. 8 – Sources, mélanges et sources séparées.

5 CONCLUSION

Nous avons présenté une étude sur la séparation aveugle de sources. Nous nous sommes intéressé au cas des mélanges convolutifs générés dans des environnements réels (salles acoustiques réverbérantes). Les résultats obtenus

dans cette étude sont concluants sur les signaux traités.

Nous avons montré qu'il est possible d'effectuer la séparation de signaux audio mélangés convolutivement, en utilisant seulement les statistiques d'ordre deux et en exploitant leur propriété de non-stationnarité. L'algorithme développé rend possible la séparation des mélanges convolutifs dans le domaine de Fourier, fréquence par fréquence. Sous l'hypothèse d'indépendance mutuelle des sources, il diagonalise conjointement l'ensemble des matrices spectrales de la même fréquence. Le critère de diagonalisation utilisé est très performant, il est basé sur l'information mutuelle Gaussienne.

Par ailleurs nous nous sommes focalisés sur la résolution de l'ambiguïté de permutation de la solution. Deux solutions ont été données. La première repose sur la continuité de la fonction de transfert des filtres de séparation. La deuxième, explore la variation de l'énergie des signaux dans le plan temps-fréquence pour construire un profil de référence à chaque signal recherché.

Pour prendre en compte la présence du bruit dans l'environnement acoustique, nous avons introduit un algorithme itératif, basé sur la maximisation de la vraisemblance. En effet, la démarche entreprise opère sur un ensemble de paramètres composé de : la matrice spectrale du bruit, la matrice de mélange et la matrice spectrale des sources. Pour implanter la démarche nous avons utilisé la technique EM, qui permet d'accomplir la tâche d'une manière simple et rapide en deux étapes.

BIBLIOGRAPHIE

- [Mansor, 2000] Mansor, A. : Blind separation for instantaneous mixture of speech signals : Algorithms and performances. *IEEE Conf.*, pages 26–32.
- [Cardoso, 1998] Cardoso, J.-F. : Blind signal separation : statistical principle. *Proceeding of the IEEE*, 9 :2009–2025.
- [Herauld et al., 1985] Herauld, J., Jutten, C., and Ans, B. : Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. *Colloque du GRETSI*, 1 :1017–1022.
- [Hyvarinen et al., 2001] Hyvarinen, A., Karhunen, J., and Oja, E. : *Independent Component Analysis*. Wiley-Interscience, New York.
- [Abed-Meraim et al., 1997] Abed-Meraim, K., Qiu, W., and Hua, Y. : Blind system identification. in *IEEE Processing*, 85, n° 8 :1310–1322.
- [Amari, 1998] Amari, S.-I. : Adaptive blind signal processing-neural network approaches. *Proceeding of the IEEE*, 86 :2026–2048.
- [Comon, 2004] Comon, P. : Blind identification and source separation in 2x3 under-determined mixtures. *IEEE Trans. on Signal Processing*, 52 :1–13.
- [Brillinger, 1975] Brillinger, D. R. : *Time series : Data analysis and theory*. Holt, Rinehart and Winston : New-York.
- [Pham, 2001] Pham, D.-T. : Joint approximate diagonalization of positive definite matrices. *SIAM J. on Matrix Anal. and Appl.*, 22(4) :1136–1152.
- [Parra and Spence, 2000] Parra, L. and Spence, C. : On line blind source separation of non stationary signals. *J. VLSI Signal Proc. Systems for Signal, Images and Video Tech.*, 26(8) :15–24.
- [Wu and Principe, 1999] Wu, H.-C. and Principe, J. C. : Simultaneous diagonalization in the frequency domain (SDIF) for source separation. In *Proceeding of ICA 1999 Conference*, pages 245–250, Aussois, France.
- [Smaragdis, 1998] Smaragdis, P. : Blind separation of convolved mixtures in the frequency domain. In *International Workshop on Independence & Artificial Neural Networks*, University of La Laguna, Tenerife, Spain.
- [Trainor et al., 2004] Trainor, L., Sonnadora, R., Wiklund, K., Bondy, J., Gupta, S., Becker, S., Bruce, I. C., and Haykin, S. :

- Development of a flexible, realistic hearing in noise test environment (R-HINT-E). *Signal Processing*, 84 :299–309.
- [Belouchrani et al., 1997] Belouchrani, A., A.-Meraim, K., Cardoso, J.-F., and Moulines, E. : A blind source separation technique using second order statistics. *IEEE Trans. on signal processing*, 45 :434–444.
- [Fleury, 2001] Fleury, G. : *Analyse spectrale ; Méthodes non-paramétriques et paramétriques*. Ellipses, Paris.
- [Max and Lacoume, 1996] Max, J. and Lacoume, J.-L. : *Méthodes et techniques de traitement du signal et applications aux mesures physiques*. Masson, Paris.
- [Westner, 1999] Westner, A. G. : *Object-Based Audio Capture : Separating Acoustically-Mixed Sounds*. PhD thesis, Massachusetts Institute of Technology, Rutgers University.
- [Mukai et al., 2004] Mukai, R., Araki, S., Sawada, H., and Makino, S. : Evaluation of separation and dereverberation performance in frequency domain blind source separation. *Acoustical Science and Technology*, 25, No.2 :119–126.
- [Mukai et al., 2001a] Mukai, R., Araki, S., and Makino, S. : Separation and dereverberation performance of frequency domain blind source separation. in *Proc. of ICA*, pages 230–235.
- [Mukai et al., 2001b] Mukai, R., Araki, S., and Makino, S. : Separation and dereverberation performance of frequency domain blind source separation for speech in a reverberant environment. in *Proc. of Eurospeech 2001*, pages 2599–2603.
- [Anemüller and Kollmeier, 2000] Anemüller, J. and Kollmeier, B. : Amplitude modulation decorrelation for convolutive blind source separation. In *Proceeding of ICA 2000 Conference*, pages 215–220, Helsinki, Finland.
- [Pham et al., 2003a] Pham, D.-T., Servière, C., and Boumaraf, H. : Blind separation of convolutive audio mixtures using nonstationarity. In *Proceeding of ICA 2003 Conference*, Nara, Japan.
- [Pham et al., 2003b] Pham, D.-T., Servière, C., and Boumaraf, H. : Blind separation of speech mixtures based on nonstationarity. In *Proceeding of ISSPA 2003 Conference*, Paris, France.
- [Boumaraf et al., 2005] Boumaraf, H., Pham, D.-T., and Servière, C. : Blind separation of convolutive mixture of speech sources. In *Proceeding of EUSIPCO 2005*, Antalya, Turkey.
- [Cardoso and Pham, 2004] Cardoso, J.-F. and Pham, D.-T. : Optimization issues in noisy gaussian ica. in *Proc. of ICA*, pages 41–48.
- [Dellaert, 2002] Dellaert, F. : The expectation maximization algorithm. Technical Report GIT-GVU-02-20, College of computing, Georgia institut of technology. <http://www.cc.gatech.edu/~dellaert/>.
- [Belouchrani, 1995] Belouchrani, A. : *Séparation autodidacte de sources : Algorithmes, Performances et Applications à des signaux expérimentaux*. PhD thesis, Ecole Nationale Supérieure des Télécommunications.
- [Belouchrani and Cardoso, 1994] Belouchrani, A. and Cardoso, J.-F. : Maximum likelihood source separation for discrete sources. in *Proc. of EUSIPCO*, 2 :768–771.
- [Belouchrani and Cardoso, 1995] Belouchrani, A. and Cardoso, J.-F. : Maximum likelihood source separation for discrete sources. in *Proc. of NOLTA*, pages 49–53.
- [Welling and Weber, 2001] Welling, M. and Weber, M. : A constrained em algorithm for principal component analysis. *Neural Computation*, 13 :677–689.
- [Ghahramani and Hinton, 1997] Ghahramani, Z. and Hinton, G. E. : The EM algorithm for mixtures of factor analyzers. Technical Report CRG-TR-96-1, Departement of computer science, University of Toronto.
- [Hartley, 1958] Hartley, H. : Maximum likelihood estimation from incomplete data. *Biometrics*, 14 :174–194.
- [Dempster et al., 1977] Dempster, A., Laird, N., and Rubin, D. : Maximum likelihood from incomplete data via the em algorithm. *Journal of Royal Statistical Society*, 39 n°1 :1–38.
- [Amari and Cichocki, 1998] Amari, S. and Cichocki, A. : A new learning algorithm for blind signal separation. *ANIPS. MIT Press, Cambridge MA*, pages 757–763.
- [Ziehe and Müller, 1998] Ziehe, A. and Müller, K.-R. : TDSEP - an efficient algorithm for blind separation using time structure. in *Proc. of ICANN*, pages 675–680.
- [Kisilev et al., 2003] Kisilev, P., Zibulevsky, M., and Zeevi, Y. Y. : A multiscale framework for blind source separation. *Journal of Machine Learning Research*, 4 :1339–1363.