# Séparation de Sources en Mélanges Convolutifs Sous-déterminés Combinant Masquage et formation de voie

Mostafa Bella<sup>1,2</sup>, Hicham Saylani<sup>1</sup>, Shahram Hosseini<sup>2</sup>, Yannick Deville<sup>2</sup>

<sup>1</sup>LETSMP, Faculté des Sciences, Université Ibn Zohr, BP 8106 Cité Dakhla, Agadir, Maroc

<sup>2</sup>IRAP, Université de Toulouse, UPS, CNRS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France \*

**Résumé** – Ce papier présente une nouvelle méthode de Séparation Aveugle de Sources (SAS) en mélanges convolutifs pouvant être sous-déterminés. Exploitant la parcimonie des signaux sources dans le domaine Temps-Fréquence (TF), cette méthode combine le masquage TF et la formation de voie (ou "beamforming"). L'utilisation du beamforming permet de réduire considérablement les artefacts qui affectent les signaux séparés par les méthodes de SAS basées uniquement sur le masquage TF. Par ailleurs, contrairement aux méthodes existantes qui utilisent le beamforming en plus du masquage TF, nous introduisons de nouveaux vecteurs directionnels dits normalisés pour générer les différents beamformers mis en jeu, et une nouvelle approche pour mieux estimer ces vecteurs, ce qui constitue un grand avantage pour notre méthode. Les tests ont montré de bonnes performances de notre méthode comparée à quelques méthodes existantes, similaires en termes d'hypothèses de travail.

Abstract - A new Blind Source Separation (BSS) method for convolutive mixtures that might be underdetermined is presented in this paper. This method, which exploits the sparsity of the source signals in the Time-Frequency (TF) domain, combines TF masking and beamforming. The artifacts that affect the signals separated by BSS methods based only on TF masking are significantly reduced by using beamforming. Moreover, unlike existing methods that combine beamforming with TF masking, we introduce new normalized directional vectors to generate the different beamformers involved, and a new way for better estimating these vectors, which is a great advantage for our method. Test results showed good performance of our method compared to some existing methods, similar in terms of working hypotheses.

## 1 Introduction

Nous nous intéressons dans ce papier aux mélanges dits convolutifs pour lesquels chaque mélange  $x_i(n)$  s'exprime en fonction de N sources  $s_i(n)$  comme suit :

$$x_i(n) = \sum_{j=1}^N h_{ij}(n) * s_j(n) = \sum_{j=1}^N s_{ij}^{\text{img}}(n), \ i \in [1, ..., P],$$
(1)

où  $h_{ij}(n)$  est la réponse impulsionnelle du filtre de mélange liant la source d'indice "j" au capteur d'indice "i" et  $s_{ij}^{img}(n)$  est appelée image temporelle de la source  $s_j(n)$  sur le capteur d'indice "i". Les méthodes de SAS basées sur le masquage TF [1, 2, 3] ont été largement utilisées récemment pour la séparation de ce type de mélanges, vu leur capacité à traiter même le cas sous-déterminé (i.e. P < N). Ces méthodes exploitent la parcimonie des sources dans le plan TF en supposant qu'elles sont *W-Disjoint Orthogonal* (WDO) dans ce plan, i.e. dans chaque point TF une seule source au plus est dominante. Ces méthodes estiment un masque de séparation propre à chaque source (qui regroupe les points TF où elle est dominante) et l'application de ce masque à chacun des mélanges dans le domaine TF permet de ne garder de ce dernier que les points TF appartenant à cette source et par la suite la séparer du reste du mélange.

Toutefois, toutes ces méthodes de SAS basées sur le masquage TF souffrent d'un problème d'artefacts au niveau des signaux séparés, lié principalement au fait que l'hypothèse de sources WDO n'est pas parfaitement vérifiée en pratique, ce qui rend leurs performances très limitées. Afin de remédier à ce problème d'artefacts, dans [4] nous avons récemment proposé une nouvelle méthode de SAS qui exploite les masques des différentes sources uniquement pour estimer les rapports entre les différents filtres de mélange dans le plan TF. Nous y montrons en effet qu'une fois ces rapports de filtres estimés, leur utilisation pour recombiner les différents mélanges dans le domaine TF nous permet au final d'estimer les images temporelles  $s_{ij}^{img}(n)$ , plutôt que leurs versions filtrées comme dans [1, 2, 5]. Néanmoins, cette méthode [4] s'applique uniquement dans le cas déterminé (i.e. P = N).

Pouvant être appliquées même dans le cas sous-déterminé, les méthodes de SAS proposées dans [6, 7, 8], quant à elles, ont adopté une autre solution qui repose sur les techniques de formation de voie ou "beamforming" [9, 10] pour fournir en sortie une estimation de chacune des sources. Ces tech-

<sup>\*</sup>Ce travail a été partiellement financé par le CNES (Centre National d'Etudes Spatiales, France) et par le CNRST (Centre National pour la Recherche Scientifique et Technique, Maroc).

niques de beamforming consistent à former un filtre spatial qui permet d'extraire le signal utile provenant d'une direction spécifique, tout en atténuant tous les autres signaux provenant des autres directions [9, 10]. Comme ces techniques ne peuvent pas être mises en œuvre en aveugle<sup>1</sup>, dans [6, 7] les auteurs commencent par estimer les différents vecteurs directionnels, en exploitant soit la différence de temps d'arrivée (TDOA) des sources entre les différents capteurs [6], soit les signaux pré-séparés par le masquage TF [7]. Cependant, l'estimation de ces vecteurs directionnels n'est pas toujours fiable, notamment lorsque les filtres de mélange sont longs, ce qui conduit à des performances très modestes du beamforming.

La nouvelle méthode que nous proposons dans ce papier utilise aussi le beamforming, mais contrairement aux méthodes existantes [6, 7], nous introduisons de nouveaux vecteurs directionnels, dits normalisés, que nous utilisons pour générer les différents beamformers. Nous montrons en effet qu'en exploitant les masques TF des sources nous obtenons une meilleure estimation de nos vecteurs directionnels normalisés auquel cas l'utilisation du beamforming nous donne de meilleures performances en termes de réduction d'artefacts au niveau des images temporelles  $s_{ij}^{img}(n)$  estimées.

## 2 Méthode proposée

La nouvelle méthode de SAS que nous proposons dans cet article procède en trois étapes qui font respectivement l'objet des sous-sections 2.1, 2.2, et 2.3 suivantes.

#### 2.1 Estimation des masques probabilistes

En transposant la relation (1) au domaine TF et en notant  $x_i(m,k)$  et  $s_j(m,k)$  la Transformée de Fourier à Court Terme (TFCT) des mélanges temporels  $x_i(n)$  et des sources  $s_j(n)$ , respectivement, nous obtenons :

$$x_i(m,k) = \sum_{j=1}^{N} h_{ij}(k) s_j(m,k) = \sum_{j=1}^{N} s_{ij}^{\text{img}}(m,k), \quad (2)$$

où m est l'indice des fenêtres temporelles au nombre de M, k est l'indice d'une bande fréquentielle parmi K bandes et  $h_{ij}(k)$  est la transformée de Fourier discrète de  $h_{ij}(n)$ .  $s_{ij}^{img}(m, k)$  est appelée image spatiale de la source  $s_j(m, k)$  sur le capteur d'indice i. En utilisant une formulation vectorielle, la relation (2) nous donne<sup>2</sup>:

$$\mathbf{x}(m) = \sum_{j=1}^{N} \mathbf{h}_j s_j(m), \qquad (3)$$

où  $\mathbf{h}_j = [h_{1j}(k), ..., h_{Pj}(k)]^T$ ,  $s_j(m) = s_j(m, k)$  et  $\mathbf{x}(m) = [x_1(m, k), ..., x_P(m, k)]^T$ . Comme dans [1, 2, 3, 5], si nous supposons que pour chaque source d'indice j il existe un

ensemble de points TF, noté  $E_j$ , sur lesquels cette source est dominante alors nous obtenons :

$$\mathbf{x}(m) = \mathbf{h}_j s_j(m), \quad \forall \, m \in E_j.$$
(4)

Comme dans [3], nous supposons que chaque vecteur de mélange  $\mathbf{x}(m)$  conditionné par  $m \in E_j$  obéit à une distribution gaussienne à valeur complexe de moyenne nulle qui est définie par :

$$p(\mathbf{x}(m)|j,\phi_j(m)\mathbf{B}_j) = \frac{1}{\pi^P \det(\phi_j(m)\mathbf{B}_j)} \cdot \exp\left(-\mathbf{x}(m)^H(\phi_j(m)\mathbf{B}_j)^{-1}\mathbf{x}(m)\right),$$
(5)

où  $\phi_j(m)$  représente la puissance spectro-temporelle (variant dans le temps) de la source  $s_j(m)$  et la matrice  $\mathbf{B}_j$ de taille  $P \times P$  modélise les caractéristiques spatiales invariantes dans le temps de cette source. D'autre part, la fonction de densité de  $\mathbf{x}(m)$  est décrite par le modèle de mélange gaussien complexe (cGMM) suivant [3, 5] :

$$p(\mathbf{x}(m)|\theta) = \sum_{j=1}^{N} \alpha_j \cdot p(\mathbf{x}(m)|j, \phi_j(m)\mathbf{B}_j), \qquad (6)$$

où  $\theta = \{\alpha_j, \mathbf{B}_j, \phi_j(m)\}_{j=1}^N$  est l'ensemble des paramètres du modèle de mélange et  $\alpha_j$  sont les rapports de mélange. Pour estimer les paramètres de cet ensemble  $\theta$  ainsi que les probabilités a posteriori à chaque point TF, nous utilisons un algorithme itératif de type *Expectation-Maximization* (*EM*). Notées  $\lambda_j(m)$ , ces probabilités a posteriori sont exprimées lors de l'étape *Expectation*, d'après le théorème de Bayes, par :  $\alpha_i : n(\mathbf{x}(m)|i|\phi_i(m)\mathbf{B}_i)$ 

$$\lambda_j(m) = \frac{\alpha_j \cdot p(\mathbf{x}(m)|j, \phi_j(m)\mathbf{B}_j)}{\sum_{j=1}^N \alpha_j \cdot p(\mathbf{x}(m)|j, \phi_j(m)\mathbf{B}_j)}.$$
 (7)

Lors de l'étape *Maximization*, les paramètres de l'ensemble  $\theta$  sont estimés en maximisant la fonction auxiliaire  $Q(\theta)$  définie par :

$$Q(\theta) = \sum_{m=0}^{M-1} \sum_{j=1}^{N} \lambda_j(m) \cdot \log\left(\alpha_j p(\mathbf{x}(m)|j, \phi_j(m)\mathbf{B}_j)\right).$$
(8)

Les règles de mise à jour des paramètres  $\alpha_j$ ,  $\phi_j(m)$  et  $\mathbf{B}_j$ sont données respectivement par les relations (9), (10) et (11) suivantes [3] :

$$\alpha_j = \frac{\sum_{m=0}^{M-1} \lambda_j(m)}{M},\tag{9}$$

$$\phi_j(m) = \frac{1}{P} \cdot \mathbf{x}^H(m) \mathbf{B}_j^{-1} \mathbf{x}(m), \tag{10}$$

$$\mathbf{B}_{j} = \frac{\sum_{m=0}^{M-1} \lambda_{j}(m) \cdot \mathbf{x}(m) \mathbf{x}^{H}(m) / \phi_{j}(m)}{\sum_{m=0}^{M-1} \lambda_{j}(m)}.$$
 (11)

Après la convergence de l'algorithme EM, la version finale de chaque probabilité a posteriori  $\lambda_j(m)$  correspond au masque probabiliste de l'une des N sources, et ce pour chaque bande fréquentielle. En faisant donc réapparaître l'indice de fréquence k, les masques désirés, notés  $M_j(m, k)$ , sont donnés par :

$$M_j(m,k) = \lambda_{\Pi_k(j)}(m,k), \ j \in [1,N],$$
(12)

<sup>1.</sup> En principe, les techniques de beamforming nécessitent d'avoir des informations sur le réseau de capteurs utilisé et sur le signal utile [9, 10].

<sup>2.</sup> Puisque chaque bande fréquentielle est traitée indépendamment lors de cette première étape, on propose dans cette section d'alléger les notations en omettant l'indice "k" correspondant.

où  $\Pi_k$  est une permutation associée à la bande fréquentielle d'indice k. Pour déterminer ces permutations  $\Pi_k$ , nous utilisons la méthode proposée dans [1] qui exploite la corrélation inter-fréquences entre les différentes séquences de probabilités a posteriori  $\lambda_j(m, k)$  comme critère pour leur classification.

#### 2.2 Estimation des vecteurs directionnels normalisés

Contrairement aux méthodes de SAS existantes [6, 7] qui cherchent à estimer chacun des vecteurs directionnels  $\mathbf{h}_j(k)$  constitué par les filtres de mélange  $h_{lj}(k), l \in [1, P]$ , nous nous intéressons dans notre méthode à l'estimation de nouveaux vecteurs directionnels dits normalisés, notés  $\mathbf{f}_{ij}(k), (i, j) \in [1, P] \times [1, N]$  et qui sont constitués plutôt par les rapports de filtres  $\frac{h_{lj}(k)}{h_{ij}(k)}$  qu'on note  $f_{ij}^{(l)}(k)$ par la suite. En effet, nous montrons que nous avons un moyen efficace pour estimer ces rapports, ce qui nous permet lors de l'étape suivante de pouvoir estimer au final les images temporelles  $s_{ij}^{img}(n)$  plutôt que leurs versions filtrées comme c'est le cas dans [1, 2, 5].

Nous savons que pour les points TF "(m,k)" qui vérifient parfaitement l'hypothèse (4) pour la source d'indice "j", on a :

$$\frac{x_l(m,k)}{x_i(m,k)} = \frac{h_{lj}(k)s_j(m,k)}{h_{ij}(k)s_j(m,k)} = \frac{h_{lj}(k)}{h_{ij}(k)} = f_{ij}^{(l)}(k).$$
(13)

Pour estimer ces rapports de filtres, nous commençons par identifier un ensemble de points TF pour chaque source d'indice "j" et en chaque bande fréquentielle d'indice "k" qui vérifient au mieux l'hypothèse de travail (4). Notés " $m_{jk}$ ", ces indices sont définis comme les indices "m" pour lesquels la valeur du masque  $M_j(m, k)$  est maximale :

$$m_{jk} = \operatorname*{argmax}_{m} M_j(m,k), \quad m \in [0, M-1].$$
(14)

Nous exploitons ensuite ces "meilleurs" points TF monosources  $(m_{jk}, k)$  pour estimer les rapports de filtres  $f_{ij}^{(l)}(k)$ ,  $l \in [1, P]$ , à travers la relation suivante, où l'opérateur de moyenne **E** [.] porte sur l'indice " $m_{jk}$ "<sup>3</sup>:

$$\mathbf{E}\left[\frac{x_{l}(m_{jk},k)}{x_{i}(m_{jk},k)}\right] = \frac{h_{lj}(k)}{h_{ij}(k)} = f_{ij}^{(l)}(k).$$
(15)

D'où une estimation des nouveaux vecteurs directionnels  $\mathbf{f}_{ij}(k)$ , qui sont idéalement définis par :

$$\mathbf{f}_{ij}(k) = [f_{ij}^{(1)}(k), ..., f_{ij}^{(P)}(k)]^T.$$
(16)

### 2.3 Beamforming MVDR

Lors de cette troisième étape de notre méthode nous utilisons un ensemble de beamformers MVDR (Minimum Variance Distortionless Response) [9] pour estimer la version finale des signaux séparés. Pour cela, nous exploitons les vecteurs directionnels normalisés  $\mathbf{f}_{ij}(k)$  en plus des vecteurs de signaux pré-séparés par masquage TF, notés  $\mathbf{y}_j(m,k)$  et définis par :

$$\mathbf{y}_{j}(m,k) = [y_{1j}(m,k), ..., y_{Pj}(m,k)]^{T}, \ j \in [1,N], \ (17)$$

où

$$y_{ij}(m,k) = x_i(m,k) \cdot M_j(m,k).$$
 (18)

Les techniques de beamforming peuvent être utilisées aussi bien dans le cas (sur-)déterminé que dans le cas sousdéterminé, mais seule une atténuation partielle des interférences est assurée dans le cas sous-déterminé [9, 10]. Ainsi, dans le cas de mélanges sous-déterminés, nous adoptons l'approche proposée dans [7] qui consiste à décomposer ces mélanges en un ensemble de mélanges déterminés puis à appliquer le beamforming. Noté  $\mathbf{z}_{j}^{(c)}(m,k)$ , le mélange d'entrée de chaque beamformer dans le cas sous-déterminé est défini comme suit [7] :

$$\mathbf{z}_{j}^{(c)}(m,k) = \mathbf{y}_{j}(m,k) + \mathbf{b}_{j}^{(c)}(m,k), \qquad (19)$$

où  $\mathbf{b}_{j}^{(c)}(m,k)$  regroupe "P-1" sorties  $\mathbf{y}_{r}(m,k), r \neq j$ , qualifiées d'interférences par rapport à  $\mathbf{y}_{j}(m,k)$ , parmi les "N-1" sorties restantes. En notant donc  $E_{jc}$  l'ensemble des indices "r" permis pour ces "P-1" interférences et C le nombre total de combinaisons possibles <sup>4</sup> on a :

$$\mathbf{b}_{j}^{(c)}(m,k) = \sum_{r \in E_{jc}} \mathbf{y}_{r}(m,k), \ c \in [1,C].$$
(20)

Noté  $\mathbf{w}_{ij}^{(c)}(k)$ , le beamformer généré pour filtrer le mélange  $\mathbf{z}_{i}^{(c)}(m,k)$  est donné par [9] :

$$\mathbf{w}_{ij}^{(c)}(k) = \frac{\mathbf{R}_{jc}^{-1}(k) \cdot \mathbf{f}_{ij}(k)}{\mathbf{f}_{ij}^{H}(k) \cdot \mathbf{R}_{jc}^{-1}(k) \cdot \mathbf{f}_{ij}(k)},$$
(21)

où  $\mathbf{R}_{ic}(k)$  est la matrice de covariance définie par :

$$\mathbf{R}_{jc}(k) = \mathbf{E}[\mathbf{b}_{j}^{(c)}(m,k) \cdot \mathbf{b}_{j}^{(c)}(m,k)^{H}].$$
 (22)

Enfin, les sorties  $\hat{s}_{ij}^{(c)}(m,k) = \mathbf{w}_{ij}^{(c)}(k)^H \cdot \mathbf{z}_j^{(c)}(m,k)$  de chaque beamformer sont moyennées pour donner l'estimation finale des images spatiales comme suit :

$$\hat{s}_{ij}^{\text{img}}(m,k) = \frac{1}{C} \cdot \sum_{c=1}^{C} \hat{s}_{ij}^{(c)}(m,k).$$
(23)

Les images temporelles  $s_{ij}^{img}(n)$  sont alors obtenues en appliquant la TFCT inverse comme suit :

$$\hat{s}_{ij}^{\text{img}}(n) = \text{ITFCT}\{\hat{s}_{ij}^{\text{img}}(m,k)\}.$$
(24)

## 3 Résultats

Nous avons effectué plusieurs tests sur différents jeux de mélanges de signaux audio fournis par la base de données  $SiSEC2011^5$  qui sont échantillonnés à  $16 \, kHz$  et d'une durée de  $10 \, s$  chacun, et ce en utilisant différents jeux de filtres. Ces filtres de mélange sont générés par la boîte

<sup>3.</sup> Il faut noter qu'il y a plusieurs bandes TF qui atteignent la valeur maximale du masque dans l'équation (14).

<sup>4.</sup> Ce nombre C de combinaisons possibles, qui n'est autre que le nombre de beamformers associé à chaque source, s'écrit :  $C = \frac{(N-1)!}{(P-1)!(N-P)!}$ .

<sup>5.</sup> http://sisec2011.wiki.irisa.fr/tiki-index.html

à outils [11] qui simule une salle acoustique de dimensions  $4.45 \, m \times 3.55 \, m \times 2.5 \, m$  et qui est caractérisée par un Temps de Réverbération variable noté  $RT_{60}$ . Dans tous les tests, les distances inter-microphones et microphonessources sont fixées respectivement à 1 m et 2 m. Pour le calcul de la TFCT, nous avons utilisé comme fenêtre de pondération une fenêtre de Hanning de longueur 2048 pour  $RT_{60} < 250 \, ms$  et 4096 pour  $RT_{60} \ge 250 \, ms$ , avec un taux de recouvrement de 75 %. Pour mesurer les performances de notre méthode nous utilisons comme critère le Rapport Signal sur Distorsion (SDR) fourni par la boite à outils BSSeval [12] et exprimé en décibel (dB). Pour la comparaison, nous utilisons des méthodes connues pour leurs bonnes performances et qui sont les mieux adaptées à nos hypothèses de travail et notre protocole de test. Il s'agit en effet des méthodes UCBSS [2] et celles proposées par Sawada et al. [1] et Cermak et al. [7] qui sont toutes applicables dans le cas sous-déterminé. Nous nous sommes alors intéressés dans nos tests à la mesure du SDR en fonction de  $RT_{60}$ , et ce pour  $RT_{60} \in \{50, 100, 150, 250, 500\}$ ms dans le cas sous-déterminé P = 2 et N = 3.

La figure 1 représente les performances obtenues. D'après cette figure, nous constatons que notre méthode est plus performante que les autres méthodes en termes de SDR quelle que soit la valeur de  $RT_{60}$ . En effet, par exemple pour  $RT_{60} = 50 ms$  notre méthode dépasse la meilleure d'entre elles d'environ 3.23 dB. Néanmoins, cette supériorité diminue légèrement lorsque  $RT_{60}$  augmente. Nous notons particulièrement que notre méthode est nettement plus performante que la méthode proposée par *Cermak et al.* [7] même si cette dernière combine elle aussi le masquage TF avec le beamforming. Cela peut s'expliquer par la pertinence de nos vecteurs directionnels normalisés comparés aux vecteurs directionnels estimés par cette méthode [7].



## 4 Conclusion

Dans ce papier, nous avons proposé une nouvelle méthode de SAS pour les mélanges convolutifs pouvant être sous-déterminés qui combine le masquage TF et le beamforming. Le beamforming est en effet introduit pour remédier au problème d'artefacts qui affectent les signaux séparés par la plupart des méthodes de SAS utilisant uniquement le masquage TF. Dans ce sens, nous avons introduit de nouveaux vecteurs directionnels, tout en proposant une technique efficace pour les estimer, ce qui garantit un beamforming plus efficace et par conséquent l'efficacité de notre nouvelle méthode. D'après les résultats des tests effectués, nos performances en termes de SDR sont nettement meilleures que celles des méthodes proposées dans [1, 2, 7]. Néanmoins, une étude de performances beaucoup plus complète est souhaitable pour confirmer ce résultat.

#### Références

- H. Sawada, S. Araki, and S. Makino, Underdetermined convolutive blind source separation via frequency binwise clustering and permutation alignment, IEEE Transactions on Audio, Speech, and Language Processing, vol. 619, pp. 516-527, March 2011.
- [2] V. G. Reju, S. N. Koh, and I. Y. Soon, Underdetermined convolutive blind source separation via time-frequency masking, IEEE Transactions on Audio, Speech, and Language Processing, vol. 18, pp. 101-116, January 2010.
- [3] N. Ito, S. Araki, and T. Nakatani, Recent advances in multichannel source separation and denoising based on source sparseness, in Audio Source Separation. Springer, 2018, pp. 279-300.
- [4] M. Bella and H. Saylani, A new sparse blind source separation method for determined linear convolutive mixtures in time-frequency domain, in Image and Signal Processing. Springer, 2020, pp. 357-366.
- [5] N. Ito, S. Araki, T. Yoshioka, and T. Nakatani, Relaxed disjointness based clustering for joint blind source separation and dereverberation, in 2014 14th IWAENC, 2014, pp. 268-272.
- [6] D.H. Tran Vu and R. Haeb-Umbach, Blind speech separation employing directional statistics in an expectation maximization framework, in 2010 IEEE ICASSP, 2010, pp. 241-244.
- [7] J. Cermak, S. Araki, H. Sawada, and S. Makino, Blind source separation based on a beamformer array and time frequency binary masking, in 2007 IEEE ICASSP, 2007, pp. 145-148.
- [8] K. Yamaoka, N. Ono and S. Makino, "Time-Frequency-Bin-Wise Linear Combination of Beamformers for Distortionless Signal Enhancement," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 3461-3475, 2021.
- [9] J. Capon, High-resolution frequency-wavenumber spectrum analysis, Proceedings of the IEEE, vol. 57, pp. 1408-1418, August 1969.
- [10] H.L. Van Trees, Adaptive beamformers, in Optimum Array Processing. John Wiley and Sons, 2002, pp. 710-916.
- [11] E.A. Lehmann and A.M. Johansson, Prediction of energy decay in room impulse responses simulated with an imagesource model, The Journal of the Acoustical Society of America, vol. 124, pp. 269-277, July 2008.
- [12] E. Vincent, R. Gribonval, and C. Fevotte, Performance measurement in blind audio source separation, IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, pp. 1462-1469, July 2006.