

Séparation aveugle de sources audio : cas des mélanges convolutifs

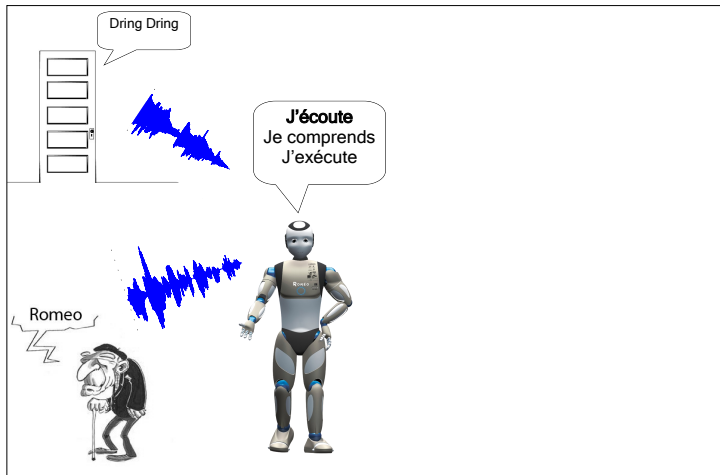
Mounira Maazaoui

Stage LIESSE

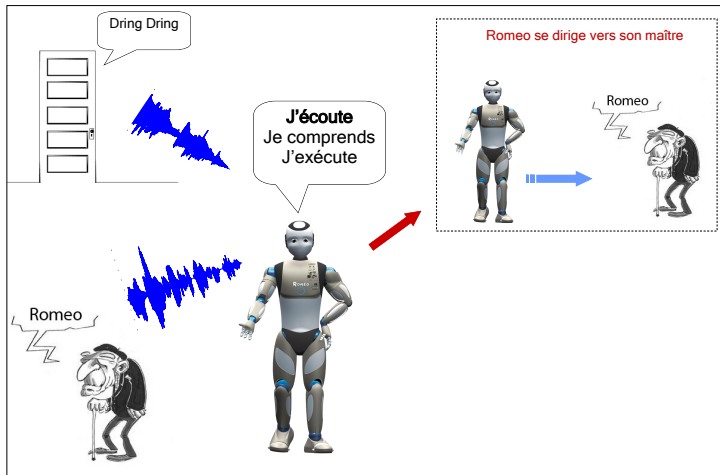
05 Novembre 2012



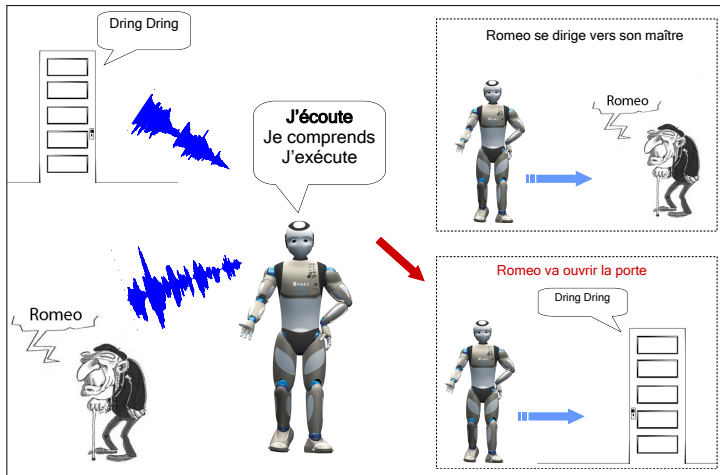
Séparation de sources : principe



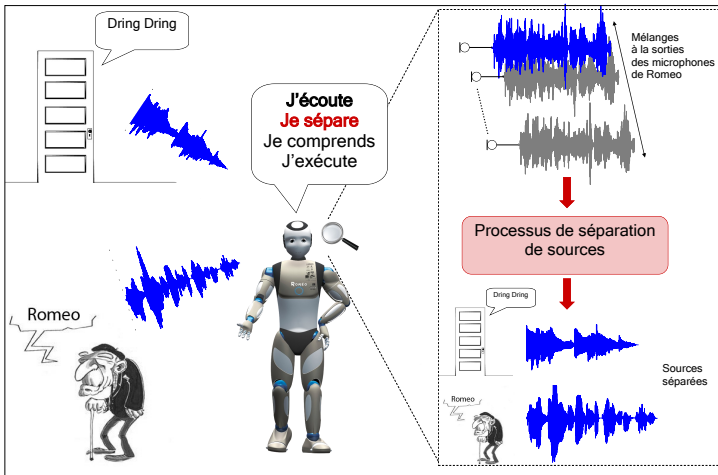
Séparation de sources : principe



Séparation de sources : principe



Séparation de sources : principe



Séparation de sources : principe

- N sources audio $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$: à estimer
- M mélanges $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T$: observations
- Modèle de mélange : $\mathbf{x}(t) = \mathcal{F}(\mathbf{s}(t))$
- Problème : à partir de $\mathbf{x}(t)$, et selon un critère de séparation, estimer $\mathbf{s}(t)$

Séparation de sources : définitions

- Séparation par réseau de microphones : 2 ou plusieurs microphones.
- Séparation de N sources à partir de M mélanges.
- Séparation **aveugle** de sources : les sources et le système de mélange ne sont pas connus *a priori*.
- Séparation de sources :
 - **déterminée** : le nombre de sources à estimer est égale au nombre de microphones $N = M$;
 - **sur-déterminée** : le nombre de sources à estimer est inférieur au nombre de microphones $N < M$;
 - **sous-déterminée** : le nombre de sources à estimer est supérieur au nombre de microphones $N > M$.

Séparation de sources : mélanges instantanés

- Mélanges *instantanés* :
le système de mélange dans le domaine temporel est une matrice $\mathbf{A} \in \mathbb{R}^{N \times M}$:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (1)$$

Pour le microphone m , le mélange est :

$$x_m(t) = \sum_{n=1}^N a_{mn} s_n(t) \quad (2)$$

Séparation de sources : mélanges convolutifs

- Mélanges **convolutifs** :

le système de mélange dans le domaine temporel est un ensemble de filtres $\mathbf{h} = [\mathbf{h}(0), \dots, \mathbf{h}(L-1)]$ où $\mathbf{h}(l) = [h_{mn}]$ $1 \leq m \leq M, 1 \leq n \leq N$:

$$\mathbf{x}(t) = \sum_{l=0}^{L-1} \mathbf{h}(l) \mathbf{s}(t-l) \quad (3)$$

Pour le microphone m , le mélange est :

$$x_m(t) = \sum_{n=1}^N \sum_{l=0}^{L-1} h_{mn}(l) s_n(t-l) \quad (4)$$

Sommaire

- 1 Etat de l'art
- 2 Séparation de mélanges convolutifs
- 3 Séparation de sources déterminée/sur-déterminée
- 4 Séparation de sources sous-déterminée
- 5 La formation de voies
- 6 Conclusion

Etat de l'art

- 1982 : formulé par Bernard Ans, Jeanny Hérault et Christian Jutten
- 1990 : le problème est généralisé pour les mélanges convolutifs
- Analyse en composantes indépendantes : introduite en 1987 par Jutten et formalisée pour les mélanges linéaires par Comon en 1991
 - minimisation des cumulants du second et troisième ordre : JADE de Cardoso et Souloumiac en 1993
 - minimisation de l'information mutuelle : Infomax de Bell et Sejnowski en 1995
- Décorrélation en utilisant les statistiques du 2nd ordre et la non-stationarité des sources : Weinstein *et al.* en 1993, Parra et Spence 2000, etc...
- Parcimonie dans le domaine temps-fréquence : masquage temps-fréquence, clustering..Ogrady *et al.* en 2005

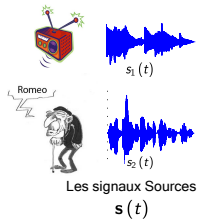
Etat de l'art

- Séparation basée sur l'analyse de scènes sonores :
 - Tordini et Piazza 2002 : utilisation du pitch dans un algorithme de séparation semi-aveugle pour séparer les sources pendant les parties voisées
 - Rutkowski et Clchocki 2001 : utilisation de la différence interaurale d'intensité (IID) et la différence interaurale de phase (IPD) pour différencier les sources
 - Perderon *et al.* 2005 : utilisation des HRTF pour poser des contraintes géométriques à la séparation

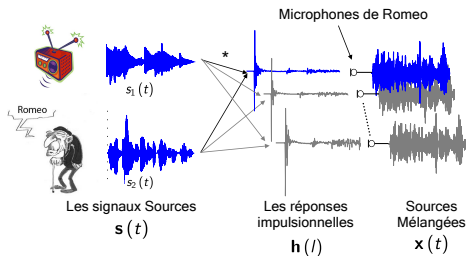
Sommaire

- 1 Etat de l'art
- 2 Séparation de mélanges convolutifs
- 3 Séparation de sources déterminée/sur-déterminée
- 4 Séparation de sources sous-déterminée
- 5 La formation de voies
- 6 Conclusion

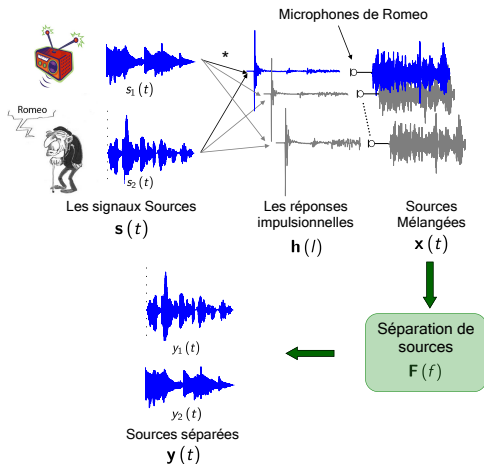
Séparation de sources dans le domaine temporel



Séparation de sources dans le domaine temporel



Séparation de sources dans le domaine temporel

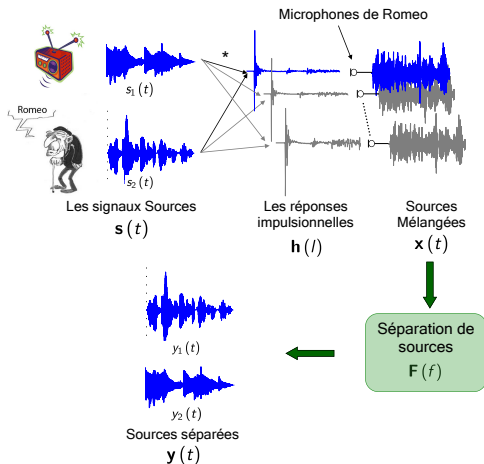


Séparation de sources dans le domaine temporel

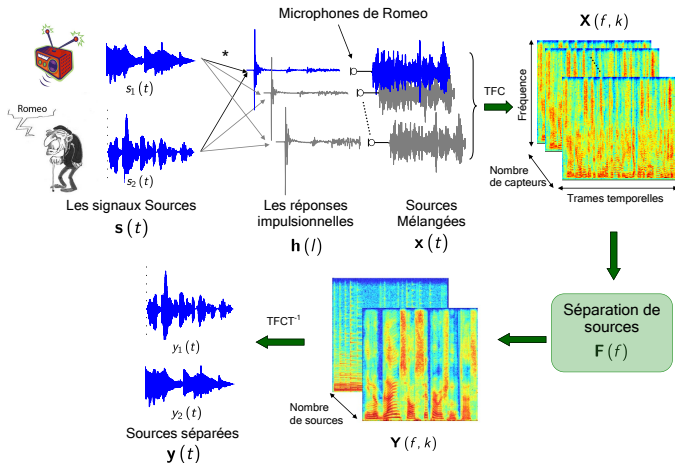
- N sources audio $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$
- M mélanges $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T = \sum_{l=0}^{L'-1} \mathbf{h}(l) \mathbf{s}(t-l)$
- Estimer des filtres de séparation $\mathbf{f} = [\mathbf{f}(0), \dots, \mathbf{f}(L'-1)]$ de longueur L' , où $\mathbf{f}(l) = [f_{ji}]_{1 \leq i \leq N, 1 \leq j \leq M}$ et telles que les sources estimées $\mathbf{y}(t)$:

$$\mathbf{y}(t) = [y_1(t), \dots, y_N(t)]^T = \sum_{l=0}^{L'-1} \mathbf{f}(l) \mathbf{x}(t-l) \quad (5)$$

Séparation de sources dans le domaine fréquentiel



Séparation de sources dans le domaine fréquentiel



Séparation de sources dans le domaine fréquentiel

- M mélanges $\mathbf{x}(t) = [x_1(t), \dots, x_M(t)]^T = \sum_{l=0}^{L-1} \mathbf{h}(l) \mathbf{s}(t-l)$
- Dans le domaine fréquentiel : $\mathbf{X}(f, k) \simeq \mathbf{H}(f) \mathbf{S}(f, k)$
 $\mathbf{X}(f, k) = [X_1(f, k), \dots, X_M(f, k)]^H$ est la TFCT de $\{\mathbf{x}(t)\}_{1 \leq t \leq T}$
 $\mathbf{S}(f, k) = [S_1(f, k), \dots, S_N(f, k)]^H$ est la TFCT de $\{\mathbf{s}(t)\}_{1 \leq t \leq T}$
- Estimer, à chaque fréquence, une matrice de séparation $\mathbf{F}(f)$ qui conduira à l'estimation des sources originales dans le domaine temps-fréquence :

$$\mathbf{Y}(f, k) = \mathbf{F}(f) \mathbf{X}(f, k) \quad (6)$$

où $\mathbf{Y}(f, k) = [Y_1(f, k), \dots, Y_N(f, k)]^H$

Problème de permutation - problème d'échelle

- Problème de *permutation* :
D'une fréquence à une autre, l'ordre des sources estimées peut changer.
 - A la fréquence f , les sources estimées sont
 $\mathbf{Y}(f, k) = [Y_1(f, k) Y_2(f, k)]^H$
 - A la fréquence $f + 1$, les sources estimées sont
 $\mathbf{Y}(f + 1, k) = [Y_2(f + 1, k) Y_1(f + 1, k)]^H$
- Problème d'*échelle* :
D'une fréquence à une autre, les sources estimées peuvent avoir des échelles différentes.

$$\mathbf{Y}(f, k) = \mathbf{F}(f) \mathbf{X}(f, k) = \mathbf{P}(f) \mathbf{D}(f) \mathbf{S}(f, k) \quad (7)$$

Sommaire

- 1 Etat de l'art
- 2 Séparation de mélanges convolutifs
- 3 Séparation de sources déterminée/sur-déterminée**
- 4 Séparation de sources sous-déterminée
- 5 La formation de voies
- 6 Conclusion

Aanalyse en composantes indépendantes

- Hypothèse : les sources sont statistiquement indépendantes et identiquement distribuées (iid) et non gaussiennes [2]
- Sources statistiquement indépendantes \implies la densité de probabilité jointe des sources $p_s(s_1(t), s_2(t), \dots, s_N(t))$ peut être factorisée en le produit des densités de probabilité marginales $\{p_i(s_i(t))\}_{1 \leq i \leq N}$:

$$p_s(s_1(t), s_2(t), \dots, s_N(t)) = p_1(s_1(t)) p_2(s_2(t)) \dots p_N(s_N(t)) \quad (8)$$

Analyse en composantes indépendantes

- Critère de séparation : minimisation de l'information mutuelle des sources
- L'information mutuelle entre les densités de probabilité jointe et marginale des sources à estimer :

$$\begin{aligned}
 I \{y_1(t), y_2(t), \dots, y_N(t)\} &= -\mathbb{E} \log \frac{p_1(y_1(t))p_2(y_2(t))\dots p_N(y_N(t))}{p_y(y_1(t), y_2(t), \dots, y_N(t))} \\
 &= \sum_{i=1}^N H \{y_i(t)\} - H \{y_1(t), \dots, y_N(t)\} \quad (9)
 \end{aligned}$$

$H \{y_1(t), \dots, y_N(t)\} = -\mathbb{E} \log p_y(y_1(t), y_2(t), \dots, y_N(t))$ est l'entropie jointe, $H \{y_i(t)\} = -\mathbb{E} \log p_i(y_i(t))$ sont des entropies marginales, $1 \leq i \leq N$

Analyse en composantes indépendantes

- Séparation de sources instantané où $\mathbf{Y}(k, f) = \mathbf{W}(f) \mathbf{X}(k, f)$, le critère de séparation est :

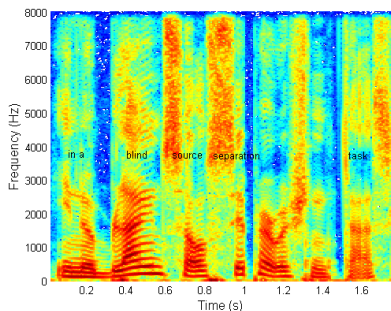
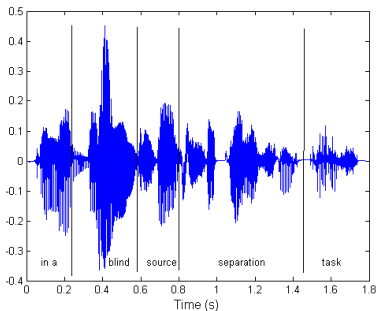
$$\psi(\mathbf{W}(f)) = \sum_{i=1}^N H(Y_i(k, f)) - \log |\det \mathbf{W}(f)| \quad (10)$$

- minimisation de l'information mutuelle des sources \Leftrightarrow maximisation de l'entropie
- Mise à jour de la matrice de séparation selon la méthode du gradient naturel [4] :

$$\mathbf{W}_{n+1}(f) = \mathbf{W}_n(f) + \mu(\mathbf{I} - \mathbf{G}_n(f)) \mathbf{W}_n(f) \quad (11)$$

où $\mathbf{G}_n(f) = \frac{1}{N_T} \sum_{k=1}^{N_T} \mathbf{f}(\mathbf{Y}_n(f, k)) \mathbf{Y}_n^H(f, k)$, $\mathbf{Y}_n(f, k) = \mathbf{W}_n(f) \mathbf{X}(f, k)$ et μ est le pas de mise à jour.

Séparation de sources basée sur la parcimonie des signaux



Minimisation de la norme l_1

- Mesurer la parcimonie d'un signal : la norme l_1 .
- La fonction de coût :

$$\psi(\mathbf{W}(f)) = \sum_{i=1}^N \sum_{k=1}^{N_T} |Y_i(f, k)| \quad (12)$$

- Estimer $\mathbf{W}(f)$ qui minimise la fonction de coût $\psi(\mathbf{W}(f))$.
- Utiliser la méthode d'optimisation du gradient naturel.
- La fonction de mise à jour de la matrice de séparation $\mathbf{W}(f)$:

$$\mathbf{W}_{t+1}(f) = \mathbf{W}_t(f) - \mu \nabla \psi(\mathbf{W}_t(f)) \mathbf{W}_t^H(f) \mathbf{W}_t(f) \quad (13)$$

Sommaire

- 1 Etat de l'art
- 2 Séparation de mélanges convolutifs
- 3 Séparation de sources déterminée/sur-déterminée
- 4 Séparation de sources sous-déterminée**
- 5 La formation de voies
- 6 Conclusion

Modèle de sources parcimonieuses : DUET

- Degenerate Unmixing Estimation Technique (DUET) [2]
- Modèle de mélange stéréo linéaire instantané : $\mathbf{X}(f, k) = \mathbf{A}\mathbf{S}(f, k)$
- Sans perte de généralité : $\mathbf{A}_{(:,n)} = [\cos \theta_n \sin \theta_n]^T, \forall n$
- Modèle de sources parcimonieuses : $\forall f, k, \exists! n_{(f,k)}$ tel que $\sigma_{n_{(f,k)}}^2(f, k) > 0$, et $\forall l \neq n_{(f,k)} \sigma_l^2(f, k) = 0$
- Si la source n est active en (f, k) , alors $\mathbf{X}(f, k) = \mathbf{a}_n s_n(f, k)$
- Estimation des paramètres θ_n et de la source active n
- Séparation de sources par masquage binaire

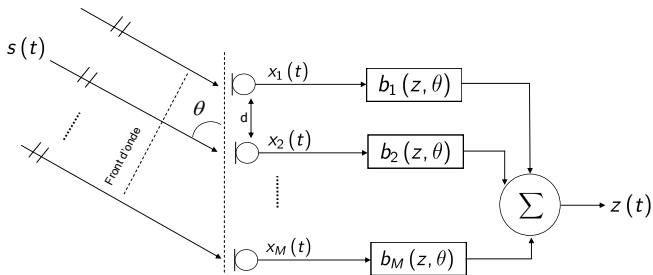
Sommaire

- 1 Etat de l'art
- 2 Séparation de mélanges convolutifs
- 3 Séparation de sources déterminée/sur-déterminée
- 4 Séparation de sources sous-déterminée
- 5 La formation de voies**
- 6 Conclusion

Principe de la formation de voies

Qu'est ce que la formation de voies ?

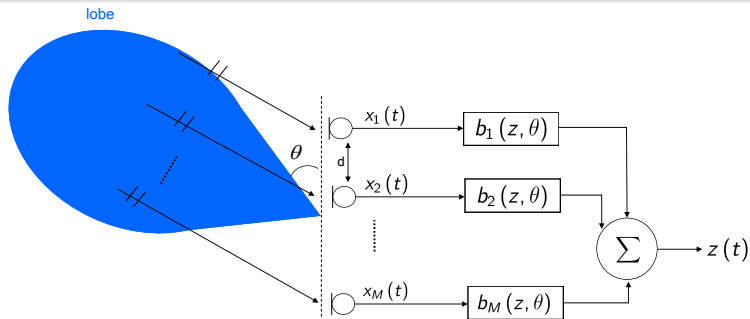
C'est un filtre spatial qui opère sur les sorties d'un réseau de capteurs dans le but de former le *diagramme de directivité* désiré [3].



Principe de la formation de voies

Qu'est ce que la formation de voies ?

C'est un filtre spatial qui opère sur les sorties d'un réseau de capteurs dans le but de former le *diagramme de directivité* désiré [3].



Principe de la formation de voies

- $z(t) = \sum_{m=1}^M \sum_{l=0}^{L-1} \underline{b}_m(l, \theta) x_m(t-l) = \sum_{l=0}^{L-1} \underline{\mathbf{b}}^T(l, \theta) \mathbf{x}(t-l)$
- Dans le domaine fréquentiel : $z(f, k) = \mathbf{b}^H(f, \theta) \mathbf{X}(f, k)$
- Réponse directionnelle : $r(f, \theta) = \mathbf{b}^H(f, \theta) \mathbf{d}(f, \theta)$

Qu'est ce qu'un vecteur directionnel ?

C'est la réponse en fréquence du réseau de capteurs.

Exemple de vecteur directionnel

Vecteur directionnel dans un champ libre d'un réseau de capteurs linéaire uniforme :

$$\mathbf{d}(f, \theta) = \left[1, e^{-j2\pi f \frac{d}{c} \sin\theta}, \dots, e^{-j2\pi f \frac{d}{c} (M-1) \sin\theta} \right] \quad (14)$$

Principe de la formation de voies

- $z(t) = \sum_{m=1}^M \sum_{l=0}^{L-1} \underline{b}_m(l, \theta) x_m(t-l) = \sum_{l=0}^{L-1} \underline{\mathbf{b}}^T(l, \theta) \mathbf{x}(t-l)$
- Dans le domaine fréquentiel : $z(f, k) = \mathbf{b}^H(f, \theta) \mathbf{X}(f, k)$
- Réponse directionnelle : $r(f, \theta) = \mathbf{b}^H(f, \theta) \mathbf{d}(f, \theta)$

Qu'est ce qu'un vecteur directionnel ?

C'est la réponse en fréquence du réseau de capteurs.

Exemple de vecteur directionnel

Vecteur directionnel dans un champ libre d'un réseau de capteurs linéaire uniforme :

$$\mathbf{d}(f, \theta) = \left[1, e^{-j2\pi f \frac{d}{c} \sin\theta}, \dots, e^{-j2\pi f \frac{d}{c} (M-1) \sin\theta} \right] \quad (14)$$

Principe de la formation de voies

- $z(t) = \sum_{m=1}^M \sum_{l=0}^{L-1} \underline{b}_m(l, \theta) x_m(t-l) = \sum_{l=0}^{L-1} \underline{\mathbf{b}}^T(l, \theta) \mathbf{x}(t-l)$
- Dans le domaine fréquentiel : $z(f, k) = \mathbf{b}^H(f, \theta) \mathbf{X}(f, k)$
- Réponse directionnelle : $r(f, \theta) = \mathbf{b}^H(f, \theta) \mathbf{d}(f, \theta)$

Qu'est ce qu'un vecteur directionnel ?

C'est la réponse en fréquence du réseau de capteurs.

Exemple de vecteur directionnel

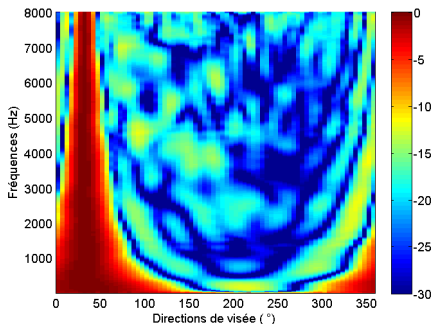
Vecteur directionnel dans un champ libre d'un réseau de capteurs linéaire uniforme :

$$\mathbf{d}(f, \theta) = \left[1, e^{-j2\pi f \frac{d}{c} \sin\theta}, \dots, e^{-j2\pi f \frac{d}{c} (M-1) \sin\theta} \right] \quad (14)$$

Principe de la formation de voies

Diagramme de directivité

$$D(f, \theta) = |r(f, \theta)|^2 = \left| \mathbf{b}^H(f, \theta) \mathbf{d}(f, \theta) \right|^2 \quad (15)$$



Principe de la formation de voies *fixe*

Définition

C'est une formation de voies *indépendante* des données observées aux capteurs.

Estimation des filtres de formation de voies fixe

Estimer $\mathbf{b}(f, \theta)$ telle que la réponse directionnelle $r(f, \theta) \approx r_d(f, \theta)$:

$$\mathbf{b}_{MC}(f, \theta) = \min_{\mathbf{b}(f, \theta)} \sum_{\theta \in \Theta} \mathcal{V}(\theta) |r(f, \theta) - r_d(f, \theta)|^2 \quad (16)$$

Exemple de réponse directionnelle désirée

$$r_d(f, \theta) = \begin{cases} 1 & \text{si } \theta_1 \leq \theta \leq \theta_2 \\ 0 & \text{sinon} \end{cases} \quad (17)$$

Principe de la formation de voies *fixe*

Définition

C'est une formation de voies *indépendante* des données observées aux capteurs.

Estimation des filtres de formation de voies fixe

Estimer $\mathbf{b}(f, \theta)$ telle que la réponse directionnelle $r(f, \theta) \approx r_d(f, \theta)$:

$$\mathbf{b}_{MC}(f, \theta) = \min_{\mathbf{b}(f, \theta)} \sum_{\theta \in \Theta} \mathcal{V}(\theta) |r(f, \theta) - r_d(f, \theta)|^2 \quad (16)$$

Exemple de réponse directionnelle désirée

$$r_d(f, \theta) = \begin{cases} 1 & \text{si } \theta_1 \leq \theta \leq \theta_2 \\ 0 & \text{sinon} \end{cases} \quad (17)$$

Principe de la formation de voies *fixe*

Définition

C'est une formation de voies *indépendante* des données observées aux capteurs.

Estimation des filtres de formation de voies fixe

Estimer $\mathbf{b}(f, \theta)$ telle que la réponse directionnelle $r(f, \theta) \approx r_d(f, \theta)$:

$$\mathbf{b}_{MC}(f, \theta) = \min_{\mathbf{b}(f, \theta)} \sum_{\theta \in \Theta} \mathcal{V}(\theta) |r(f, \theta) - r_d(f, \theta)|^2 \quad (16)$$

Exemple de réponse directionnelle désirée

$$r_d(f, \theta) = \begin{cases} 1 & \text{si } \theta_1 \leq \theta \leq \theta_2 \\ 0 & \text{sinon} \end{cases} \quad (17)$$

Formation de voies fixe vers des directions de visée fixe

- Le vecteur directionnel correspondant :

$$\mathbf{d}(f, \theta) = [h_1(f, \theta), \dots, h_M(f, \theta)]^T \quad (18)$$

- La réponse directionnelle désirée :

$$r_d(f, \theta) = \begin{cases} 1 & \text{si } \theta = \theta_i \\ 0 & \text{sinon} \end{cases} \quad (19)$$

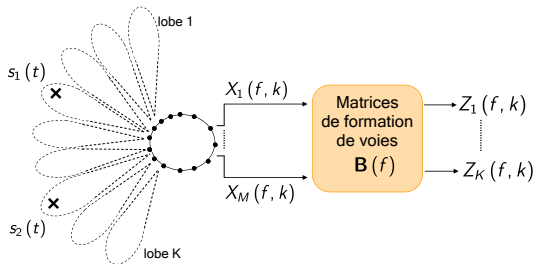
- Le filtre de formation de voies selon la méthode des moindres carrés :

$$\mathbf{b}(f, \theta_i) = \mathbf{R}_{dd}^{-1}(f) \mathbf{d}(f, \theta_i) \quad \text{où} \quad \mathbf{R}_{dd}(f) = \frac{1}{N_S} \sum_{\theta \in \Theta} \mathbf{d}(f, \theta) \mathbf{d}^H(f, \theta) \quad (20)$$

- Pour K directions de visée $\theta_1, \dots, \theta_K$:

$$\mathbf{B}(f) = [\mathbf{b}(f, \theta_1), \dots, \mathbf{b}(f, \theta_K)]^T \quad (21)$$

Formation de voies fixe vers des directions de visée fixe



BF-fixed

- Directions de visée = K directions choisies *a priori* $\theta_1, \dots, \theta_K$
- La matrice de formation de voies : $\mathbf{B}(f) = [\mathbf{b}(f, \theta_1), \dots, \mathbf{b}(f, \theta_K)]^T$
- Séparation de N sources à partir de K mélanges $\mathbf{Z}(f, k)$

Sommaire

- 1 Etat de l'art
- 2 Séparation de mélanges convolutifs
- 3 Séparation de sources déterminée/sur-déterminée
- 4 Séparation de sources sous-déterminée
- 5 La formation de voies
- 6 Conclusion**

Conclusion

- Problème de séparation de sources dans un milieu réel : modèle de mélanges convolutifs
- Séparation de mélanges convolutifs dans le domaine fréquentiel :
 - estimer les sources à chaque fréquence
 - estimer des matrices de séparation : résoudre des mélanges instantanés
- Analyse en composantes indépendantes : un cadre général pour la séparation de sources basé sur l'hypothèse d'indépendance des sources, plusieurs critères de séparation dont la minimisation de l'information mutuelle.
- La formation de voies : méthode de séparation basée sur l'information *spatiale* des sources → séparation géométrique des sources

Bibliographie

- [1] Independent Component Analysis : a new concept ?, Signal Processing, 1994
- [2] Scott Rickard. The DUET Blind Source Separation Algorithm, pages 217–241. Springer Netherlands, 2007.
- [3] B.D. Van Veen et K.M. Buckley : Beamforming : a versatile approach to spatial filtering. IEEE ASSP Magazine, pages 4–24, 1988.
- [4] Amari, S., Cichocki, A. & Yang, H. H.A, New Learning Algorithm for Blind Signal Separation, Advances in Neural Information, 1996. Processing Systems