

Séparation de sources audio

M2 Mathématiques / Vision / Apprentissage
Analyse des signaux audiofréquences

PAGE 1 / 46



Roland Badeau



Partie I

Introduction

Page 2 / 46



Roland Badeau



Introduction

- Séparation de sources
 - Art d'estimer des signaux "sources", supposés indépendants, à partir de l'observation d'un ou plusieurs "mélanges" de ces sources
- Exemples applicatifs :
 - Débruitage (cocktail party, suppression du vuvuzela, karaoké)
 - Séparation des instruments dans une musique polyphonique
 - Remixage, transformations, respatialisation

PAGE 3 / 46



Roland Badeau



Typologie des modèles de mélanges

- Définition du problème
 - Observations : M mélanges $x_m(t)$, concaténés dans un vecteur $\mathbf{x}(t)$
 - Inconnues : K sources $s_k(t)$, concaténées dans un vecteur $\mathbf{s}(t)$
 - Modèle général de mélange : $\mathbf{x}(t) = \mathcal{A}_t(\mathbf{s}(:))$
- Stationnarité : \mathcal{A} ne dépend pas de t ($\mathcal{A}_t = \mathcal{A}$)
- Linéarité : \mathcal{A} est une application linéaire
- Mémoire :
 - Mélanges convolutifs
 - Mélanges instantanés : $\mathbf{x}(t) = \mathbf{A} \mathbf{s}(t)$
 - \mathcal{A} est définie par la "matrice de mélange" \mathbf{A} (de dimension $M \times K$)
- Inversibilité :
 - Mélanges déterminés : $M = K$
 - Mélanges sur-déterminés : $M > K$
 - Mélanges sous-déterminés : $M < K$

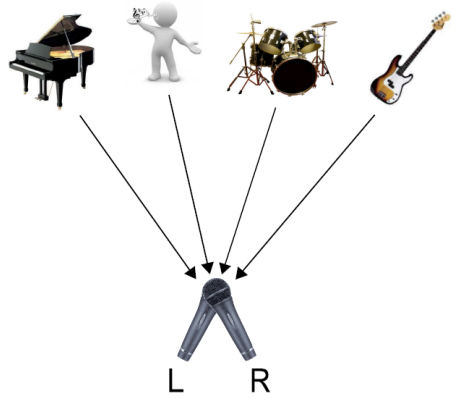
PAGE 4 / 46



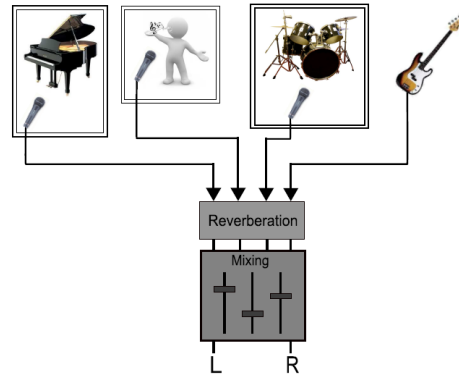
Roland Badeau



Mélanges linéaires instantanés

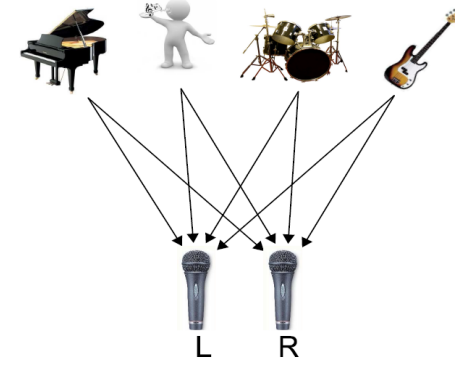


(a) Configuration stéréo XY.

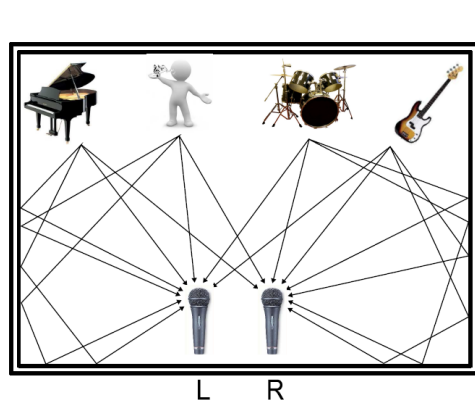


(b) Injection directe sur la table de mixage.

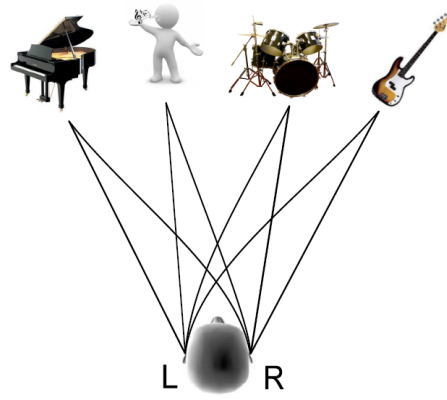
Mélanges linéaires anéchoïques



Mélanges linéaires convolutifs



(a) Mélange convolutif.



(b) Mélange binaural.



Partie II

Rappels mathématiques

- Notation : $\phi[\mathbf{x}]$ désigne une fonction de $p(\mathbf{x})$
- Vecteur moyenne : $\mu_x = \mathbb{E}[\mathbf{x}]$
- Matrice de covariance : $\mathbf{R}_{xx} = \mathbb{E}[(\mathbf{x} - \mu_x)(\mathbf{x} - \mu_x)^T]$
- Fonction caractéristique :

$$\phi_x(\mathbf{f}) = \mathbb{E}[e^{-2i\pi\mathbf{f}^T\mathbf{x}}] = \int_{\mathbb{R}} p(\mathbf{x}) e^{-2i\pi\mathbf{f}^T\mathbf{x}} d\mathbf{x}$$

- Distribution de probabilité : $p(\mathbf{x}) = \int_{\mathbb{R}} \phi_x(\mathbf{f}) e^{+2i\pi\mathbf{f}^T\mathbf{x}} d\mathbf{f}$
- Cumulants :

- Définition : $\ln(\phi_x(\mathbf{f})) = \sum_{n=1}^{+\infty} \frac{(-2i\pi)^n}{n!} \sum_{k_1=1}^K \sum_{k_n=1}^K \kappa_{k_1 \dots k_n}^n[\mathbf{x}] f_{k_1} \dots f_{k_n}$
- $\kappa^n[\mathbf{x}]$ est un tenseur d'ordre n
- $\kappa^1[\mathbf{x}]$ est le vecteur moyenne, $\kappa^2[\mathbf{x}]$ est la matrice de covariance
- Si $p(\mathbf{x})$ est symétrique ($p(-\mathbf{x}) = p(\mathbf{x})$), $\kappa^n[\mathbf{x}] = 0$ pour tout n impair
- $\kappa_{k,k,k,k}^4[\mathbf{x}]$ est appelé "kurtosis"

- La distribution gaussienne est celle pour laquelle tous les cumulants d'ordre $n > 2$ sont nuls
- Fonction caractéristique

$$\phi_x(\mathbf{f}) = \exp(-2i\pi\mathbf{f}^T\mu_x - 2\pi^2\mathbf{f}^T\mathbf{R}_{xx}\mathbf{f})$$

- Distribution de probabilités (définie si \mathbf{R}_{xx} est inversible)

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{K}{2}} \det(\mathbf{R}_{xx})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_x)^T \mathbf{R}_{xx}^{-1} (\mathbf{x} - \mu_x)\right)$$

- Définition : les cumulants d'ordre 1 et 2 sont invariants par translation temporelle
- Matrices de covariances de 2 processus SSL centrés $\mathbf{x}(t)$ et $\mathbf{y}(t)$:
 - Définition : $\mathbf{R}_{xy}(\tau) = \mathbb{E}[\mathbf{x}(t+\tau)\mathbf{y}(t)^T]$
 - Propriété : $\mathbf{R}_{xx}(0)$ est symétrique hermitienne positive
- Matrices de DSP d'un processus SSL $\mathbf{x}(t)$:
 - Définition : $\mathbf{S}_{xx}(\nu) = \sum_{\tau \in \mathbb{Z}} \mathbf{R}_{xx}(\tau) e^{-2i\pi\nu\tau}$
 - Propriété : $\forall \nu$, $\mathbf{S}_{xx}(\nu)$ est symétrique hermitienne positive

- Entropie de Shannon
 - Définition : $H[\mathbf{x}] = -\mathbb{E}[\ln(p(\mathbf{x}))]$
 - $H[\mathbf{x}]$ n'est pas nécessairement positive pour une v.a. continue
- Divergence de Kullback-Leibler
 - $D_{KL}(p||q) = \int p(\mathbf{x}) \ln\left(\frac{p(\mathbf{x})}{q(\mathbf{x})}\right) d\mathbf{x}$
 - Propriété : $D_{KL}(p||q) \geq 0$, $D_{KL}(p||q) = 0$ ssi $p = q$
- Information mutuelle
 - Définition : $I[\mathbf{x}] = \mathbb{E}\left[\ln\left(\frac{p(\mathbf{x})}{p(x_1)\dots p(x_K)}\right)\right] = D_{KL}(p(\mathbf{x})||p(x_1)\dots p(x_K))$
 - Propriété : $I[\mathbf{x}] = 0$ ssi $x_1 \dots x_K$ sont mutuellement indépendantes
 - Lien avec l'entropie : $I[\mathbf{x}] = \sum_{k=1}^K H[x_k] - H[\mathbf{x}]$



Partie III

Mélanges linéaires instantanés



Séparation aveugle de sources

- Modèle d'observation *Blind Source Separation* (BSS) :
 - $\forall t, \mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$ où $\mathbf{A} \in \mathbb{R}^{M \times K}$ est appelée "matrice de mélange"
 - Distribution IID des sources : $p(\{s_k(t)\}_{k,t}) = \prod_{k=1}^K \prod_{t=1}^T p_k(s_k(t))$
- Problème : estimer \mathbf{A} et les sources $\mathbf{s}(t)$ connaissant $\mathbf{x}(t)$
- Définition : Matrice non mélangeante
 - une matrice \mathbf{C} de dimension $K \times K$ est non mélangeante ssi elle a un unique coefficient non nul dans chaque ligne et chaque colonne
- Si on pose $\tilde{\mathbf{s}}(t) = \mathbf{C}\mathbf{s}(t)$ et $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{C}^{-1}$, alors $\mathbf{x}(t) = \tilde{\mathbf{A}}\tilde{\mathbf{s}}(t)$ est une autre décomposition admissible des observations
 - On ne peut retrouver les sources qu'à une permutation et un coefficient multiplicatif près



Séparation linéaire des sources

- Soit $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$, où $\mathbf{B} \in \mathbb{R}^{K \times M}$ est dite "matrice de séparation"
- Si la matrice \mathbf{A} est connue :
 - Si \mathbf{A} est de rang K , on obtient $\mathbf{y}(t) = \mathbf{s}(t)$ en posant :
 - $\mathbf{B} = \mathbf{A}^{-1}$ dans le cas déterminé ($M = K$)
 - $\mathbf{B} = \mathbf{A}^\dagger$ dans le cas sur-déterminé ($M > K$)
 - la pseudo-inverse $\mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ est telle que $\mathbf{A}^\dagger \mathbf{A} = \mathbf{I}_K$
 - Dans le cas sous-déterminé ($M < K$), la séparation est impossible



Analyse en composantes indépendantes

- Si la matrice \mathbf{A} est inconnue :
 - On cherche une matrice \mathbf{B} qui rend les y_k indépendants (ACI)
 - On obtient alors la relation $\mathbf{y}(t) = \mathbf{C}\mathbf{s}(t)$, où $\mathbf{C} = \mathbf{B}\mathbf{A}$
 - Le problème est résolu si la matrice \mathbf{C} est non mélangeante



- On suppose à présent que les sources sont centrées : $\mathbb{E}[\mathbf{s}(t)] = \mathbf{0}$ et que le mélange est (sur-)déterminé
- Problème canonique : on peut supposer sans perte de généralité que $\mathbf{s}(t)$ est spatialement blanc ($\mathbf{R}_{ss} = \mathbb{E}[\mathbf{s}(t)\mathbf{s}(t)^T] = \mathbf{I}_K$)
- Alors $\mathbf{R}_{xx} = \mathbf{A}\mathbf{R}_{ss}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T$: \mathbf{A} est une racine carrée de \mathbf{R}_{xx}
- On cherche d'abord à blanchir (décorrélérer) le mélange :
 - \mathbf{R}_{xx} est diagonalisable dans une base orthonormée : $\mathbf{R}_{xx} = \mathbf{Q}\mathbf{\Sigma}^2\mathbf{Q}^T$ où $\mathbf{\Sigma} = \text{diag}(\sqrt{\lambda_1} \dots \sqrt{\lambda_M})$ avec $\lambda_1 \geq \lambda_K > \lambda_{K+1} = \dots = \lambda_M = 0$ (le rang de \mathbf{R}_{xx} est égal à K)
 - On pose $\mathbf{S} = \mathbf{Q}_{(:,1:K)}\mathbf{\Sigma}_{(1:K,1:K)} \in \mathbb{R}^{M \times K}$
 - \mathbf{S} est une racine carrée matricielle de \mathbf{R}_{xx} : $\mathbf{R}_{xx} = \mathbf{S}\mathbf{S}^T$
 - On pose $\mathbf{W} = \mathbf{S}^\dagger$ et $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$
 - Alors $\mathbf{z}(t)$ est blanc ($\mathbb{E}[\mathbf{z}(t)] = \mathbf{0}$ et $\mathbf{R}_{zz} = \mathbf{W}\mathbf{R}_{xx}\mathbf{W}^T = \mathbf{I}$)

- On en déduit sans perte de généralité que $\mathbf{U} \triangleq \mathbf{W}\mathbf{A}$ est une matrice de rotation ($\mathbf{U}\mathbf{U}^T = \mathbf{I}$).
- Alors $\mathbf{y}(t) = \mathbf{U}^T\mathbf{z}(t) = \mathbf{U}^T\mathbf{W}\mathbf{x}(t) = (\mathbf{W}\mathbf{A})^{-1}(\mathbf{W}\mathbf{A})\mathbf{s}(t) = \mathbf{s}(t)$.
- On peut donc poser $\mathbf{B} = \mathbf{U}^T\mathbf{W}$ où \mathbf{U} est une rotation.

- On peut estimer \mathbf{R}_{xx} à partir des observations et en déduire \mathbf{W}
- La blancheur (cumulants d'ordre 2) détermine \mathbf{W} et laisse \mathbf{U} inconnue.
- Si les sources sont gaussiennes, les z_k sont indépendants et on ne peut pas déterminer \mathbf{U} .
- Pour déterminer la rotation \mathbf{U} , il faut exploiter la non gaussianité des sources et caractériser la propriété d'indépendance en utilisant les cumulants d'ordre supérieur à 2.

- Définition : ϕ est une "fonction de contraste" ssi $\phi[\mathbf{C}\mathbf{s}(t)] \geq \phi[\mathbf{s}(t)] \forall \mathbf{C}$ et si $\phi[\mathbf{C}\mathbf{s}(t)] = \phi[\mathbf{s}(t)] \Leftrightarrow \mathbf{C}$ est non mélangeante.
- La séparation se fait en minimisant $\phi[\mathbf{y}(t) = \mathbf{C}\mathbf{s}(t)]$ par rapport à \mathbf{U} (ou \mathbf{B})
- Fonction de contraste "canonique" : $\phi_{IM}[\mathbf{y}(t)] = I[\mathbf{y}(t)]$
- Contrastes orthogonaux : à minimiser sous la contrainte $\mathbb{E}[\mathbf{y}(t)\mathbf{y}(t)^T] = \mathbf{I}$. Par exemple, $\phi_{IM}^\circ[\mathbf{y}(t)] = \sum_{k=1}^K H(y_k(t))$
- Approximation à l'ordre 4 de ϕ_{IM}° : $\phi_{ICA}^\circ[\mathbf{y}(t)] = \sum_{ijkl \neq iiii} \kappa_{ijkl}^4[\mathbf{y}(t)]$
- Algorithmes de descente de la fonction ϕ par rapport à \mathbf{B} ou \mathbf{U} :
 - Algorithme de gradient appliqué à la matrice \mathbf{B}
 - Paramétrisation de \mathbf{U} avec des rotations de Givens et relaxation

1. Estimation de la matrice de covariance $\hat{\mathbf{R}}_{xx}$
2. Diagonalisation de $\hat{\mathbf{R}}_{xx}$: $\hat{\mathbf{R}}_{xx} = \hat{\mathbf{Q}}\hat{\mathbf{\Sigma}}^2\hat{\mathbf{Q}}^T$ où $\hat{\mathbf{\Sigma}} = \text{diag}(\sqrt{\hat{\lambda}_1} \dots \sqrt{\hat{\lambda}_M})$ avec $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_M \geq 0$
3. Calcul de $\mathbf{S} = \hat{\mathbf{Q}}_{(:,1:K)}\hat{\mathbf{\Sigma}}_{(1:K,1:K)}$
4. Calcul du blanchisseur $\mathbf{W} = \mathbf{S}^\dagger$
5. Blanchiment des données : $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$
6. Estimation de \mathbf{U} par minimisation de la fonction de contraste ϕ°
7. Estimation des signaux sources par $\mathbf{y}(t) = \mathbf{U}^T\mathbf{z}(t)$

- Modèle : $\mathbb{E}(\mathbf{s}(t)) = \mathbf{0}$, $\mathbf{R}_{ss}(\tau) = \mathbb{E}(\mathbf{s}(t+\tau)\mathbf{s}(t)^T) = \text{diag}(r_{s_k}(\tau))$
- Problème canonique : on suppose que $\mathbf{R}_{ss}(0) = \mathbf{I}$
- On cherche d'abord à blanchir spatialement le mélange :
 - Soit \mathbf{S} une racine carrée matricielle de $\mathbf{R}_{xx}(0)$
 - On pose $\mathbf{W} = \mathbf{S}^\dagger$ et $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$
- Comme $\mathbf{R}_{xx}(0) = \mathbf{A}\mathbf{A}^T$, $\mathbf{U} \triangleq \mathbf{W}\mathbf{A}$ est une matrice de rotation
- Or $\forall \tau \in \mathbb{Z}$, $\mathbf{R}_{zz}(\tau) = \mathbf{U}\mathbf{R}_{ss}(\tau)\mathbf{U}^T$
- La diagonalisation conjointe des matrices $\mathbf{R}_{zz}(\tau)$ pour plusieurs valeurs de τ permet d'identifier la rotation \mathbf{U}

- Théorème d'unicité :
 - Soit un ensemble de matrices $\mathbf{R}_{zz}(\tau)$ de dimension $K \times K$ et de la forme $\mathbf{R}_{zz}(\tau) = \mathbf{U}\mathbf{R}_{ss}(\tau)\mathbf{U}^T$ avec \mathbf{U} unitaire et $\mathbf{R}_{ss}(\tau) = \text{diag}(r_{s_k}(\tau))$. Alors \mathbf{U} est unique (à une matrice non mélangeante près) ssi $\forall 1 \leq k \neq l \leq K$, il existe τ tel que $r_{s_k}(\tau) \neq r_{s_l}(\tau)$
- Méthodes de diagonalisation conjointe : on minimise le critère

$$J(\mathbf{U}) = \sum_{\tau} \|\mathbf{U}^T\mathbf{R}_{zz}(\tau)\mathbf{U} - \text{diag}(\mathbf{U}^T\mathbf{R}_{zz}(\tau)\mathbf{U})\|_F^2$$
 - Paramétrisation de \mathbf{U} avec des rotations de Givens et algorithme de relaxation

- *Second Order Blind Identification (SOBI)*
 1. Estimation et diagonalisation de $\hat{\mathbf{R}}_{xx}(0)$: $\hat{\mathbf{R}}_{xx}(0) = \hat{\mathbf{Q}}\hat{\mathbf{\Sigma}}^2\hat{\mathbf{Q}}^T$ où $\hat{\mathbf{\Sigma}} = \text{diag}(\sqrt{\hat{\lambda}_1} \dots \sqrt{\hat{\lambda}_M})$ avec $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_M \geq 0$
 2. Calcul de $\mathbf{S} = \hat{\mathbf{Q}}_{(:,1:K)}\hat{\mathbf{\Sigma}}_{(1:K,1:K)}$
 3. Calcul du blanchisseur $\mathbf{W} = \mathbf{S}^\dagger$
 4. Blanchiment des données : $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$
 5. Estimation des matrices de covariance $\hat{\mathbf{R}}_{zz}(\tau)$ à différents retards τ
 6. Diagonalisation conjointe approchée des matrices $\hat{\mathbf{R}}_{zz}(\tau)$ dans une base \mathbf{U}
 7. Estimation des signaux sources par $\mathbf{y}(t) = \mathbf{U}^T\mathbf{z}(t)$

- Modèle : $\mathbb{E}(\mathbf{s}(t)) = \mathbf{0}$, $\mathbf{R}'_{ss}(t) \triangleq \mathbb{E}(\mathbf{s}(t)\mathbf{s}(t)^T) = \text{diag}(\sigma_k^2(t))$
- Alors $\forall t \in \mathbb{Z}$, $\mathbf{R}'_{xx}(t) = \mathbf{A}\mathbf{R}'_{ss}(t)\mathbf{A}^T$
- Méthodes de diagonalisation conjointe : on minimise le critère

$$J(\mathbf{B}) = \sum_t \|\mathbf{B}\mathbf{R}'_{xx}(t)\mathbf{B}^T - \text{diag}(\mathbf{B}\mathbf{R}'_{xx}(t)\mathbf{B}^T)\|_F^2$$
 - Algorithme de descente de gradient appliqué à la matrice \mathbf{B}
 - Dans le cas surdéterminé, il faut contraindre \mathbf{B} à engendrer le sous-espace principal de toutes les matrices $\mathbf{R}'_{xx}(t)$
- Variante de l'algorithme SOBI :
 1. Segmentation des signaux sources et estimation des matrices de covariance $\mathbf{R}'_{xx}(t)$ sur des fenêtres centrées à différents instants t
 2. Diagonalisation conjointe des matrices $\mathbf{R}'_{xx}(t)$ dans une base \mathbf{B}
 3. Estimation des signaux sources par $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$

- L'utilisation des cumulants d'ordre supérieur n'est nécessaire que pour le modèle de sources IID non gaussiennes
- Les statistiques du second ordre suffisent pour des sources :
 - stationnaires mais non IID (\rightarrow dynamique spectrale)
 - non stationnaires (\rightarrow dynamique temporelle)
- On retiendra que les outils classiques (basés sur les statistiques d'ordre 2) sont appropriés pour la séparation aveugle de sources indépendantes (et éventuellement gaussiennes), à condition de tenir compte de la dynamique spectrale / temporelle des sources.



Partie IV

Méthodes temps-fréquence

Représentations temps-fréquence

- Motivations
 - Les dynamiques spectrale et temporelle sont mises en évidence par une représentation temps-fréquence (TF) des signaux
 - Les représentations TF sont appropriées pour traiter les cas des mélanges convolutifs et/ou sous-déterminés
- Utilisation d'un banc de filtres (exemples : TFCT, MDCT) :
 - Découpage en F sous-bandes et décimation d'un facteur $T \leq F$
 - Filtres d'analyse h_f et de synthèse g_f
 - Représentation TF des sources : $s_k(f, n) = (h_f * s_k)(nT)$
 - Représentation TF des mélanges : $x_m(f, n) = (h_f * x_m)(nT)$
 - Reconstruction parfaite : $s_k(t) = \sum_{f=1}^F g_f(t - nT)s_k(f, n)$
- Alors $\forall f, n$, $\mathbf{x}(f, n) = \mathbf{A}\mathbf{s}(f, n)$ (même mélange linéaire instantané)

- Hypothèse : sources gaussiennes indépendantes et centrées
 - Modèle de **sources instationnaires** :
 - si les trames temporelles n_1 et n_2 sont disjointes, alors $s_k(\cdot, n_1)$ et $s_k(\cdot, n_2)$ sont indépendants et de variances distinctes
 - Modèle de **sources stationnaires cohérentes** :
 - si les sous-bandes f_1 et f_2 sont disjointes ($h_{f_1} * \tilde{h}_{f_2} = 0$), alors $s_k(f_1, \cdot)$ et $s_k(f_2, \cdot)$ sont stationnaires, indépendants et de variances distinctes $\sigma_k^2(f_1) = (h_{f_1} * \tilde{h}_{f_1} * r_{s_k})(0)$ et $\sigma_k^2(f_2) = (h_{f_2} * \tilde{h}_{f_2} * r_{s_k})(0)$
 - Nouveau modèle de sources :
 - tous les $s_k(f, n)$ sont indépendants pour tous n et f , de variances distinctes $\sigma_k^2(f, n)$ (\Rightarrow dynamique temps-fréquence)



- Séparation par diagonalisation conjointe de matrices :
 - Soient $\mathbf{R}_{ss}(f, n) = \mathbb{E}[\mathbf{s}(f, n)\mathbf{s}(f, n)^T]$ et $\mathbf{R}_{xx}(f, n) = \mathbb{E}[\mathbf{x}(f, n)\mathbf{x}(f, n)^T]$
 - Alors $\mathbf{R}_{xx}(f, n) = \mathbf{A}\mathbf{R}_{ss}(f, n)\mathbf{A}^T$ où $\mathbf{R}_{ss}(f, n) = \text{diag}(\sigma_k^2(f, n))$
- Variante de l'algorithme SOBI :
 1. Analyse TF des mélanges : $u_k(f, \cdot) = h_f * x_k$ et $x_k(f, n) = u_k(f, nT)$
 2. Estimation des matrices de covariance $\mathbf{R}_{xx}(f, n)$:

$$\hat{\mathbf{R}}_{xx}(f, n) = \sum_{t \in \mathbb{Z}} w(t - nT) \mathbf{u}(f, t)\mathbf{u}(f, t)^T$$
 où $w \geq 0$ est une fenêtre
 3. Diagonalisation conjointe des matrices $\hat{\mathbf{R}}_{xx}(f, n)$ dans une base \mathbf{B}
 4. Estimation des signaux sources par $\mathbf{y}(f, n) = \mathbf{B}\mathbf{x}(f, n)$
 5. Synthèse TF des sources : $y_k(t) = \sum_{f=1}^F g_f(t - nT)y_k(f, n)$



Partie V

Mélanges convolutifs

Décomposition du problème de séparation

- Modèle de mélange instantané inadapté aux mélanges acoustiques réels
- Soit $\mathbf{x}_k(f, n) \in \mathbb{R}^M$ l'**image** de la source $s_k(f, n)$
 - signal vectoriel capté si seule la source $s_k(f, n)$ était active
- Modèle de mélange : $\mathbf{x}(f, n) = \sum_{k=1}^K \mathbf{x}_k(f, n)$
- Décomposition du problème de séparation de sources
 - **séparation** : estimer $\mathbf{x}_k(f, n)$ à partir du mélange $\mathbf{x}(f, n)$
 - **déconvolution** : estimer $s_k(f, n)$ à partir de $\mathbf{x}_k(f, n)$



- Modèle de mélange : $x_m(t) = \sum_{k=1}^K (a_{mk} * s_k)(t)$
- Hypothèses :
 - le banc de filtres correspond à une TFCT
 - la RI de a_{mk} est courte devant la longueur des fenêtres
 - $\forall m, k, f, a_{mk}(f)$ varie lentement devant $h_f(\nu)$
 - $\Rightarrow (h_f * a_{mk})(t) \approx a_{mk}(f) h_f(t)$
- Approximation du modèle de mélange convolutif :

$$x_m(f, n) = \sum_{k=1}^K a_{mk}(f) s_k(f, n)$$
- Écriture matricielle : $\mathbf{x}(f, n) = \mathbf{A}(f) \mathbf{s}(f, n)$
 - F modèles de mélanges instantanés dans chaque sous-bande
 - \Rightarrow on peut utiliser n'importe quelle méthode d'ACI par sous-bande



- Soit $\mathbf{y}(f, n) = \mathbf{B}(f) \mathbf{x}(f, n)$, où $\mathbf{B}(f) \in \mathbb{C}^{K \times M}$
- Si les matrices $\mathbf{A}(f)$ sont connues :
 - Si $\mathbf{A}(f)$ est de rang K , on obtient $\mathbf{y}(f, n) = \mathbf{s}(f, n)$ en posant :
 - $\mathbf{B}(f) = \mathbf{A}(f)^{-1}$ dans le cas déterminé ($M = K$)
 - $\mathbf{B}(f) = \mathbf{A}(f)^\dagger$ dans le cas sur-déterminé ($M > K$)
 - Dans le cas sous-déterminé ($M < K$), la séparation reste impossible
- Si la matrice $\mathbf{A}(f)$ est inconnue :
 - On cherche $\mathbf{B}(f)$ qui rend les $y_k(f, n)$ indépendants (ACI)
 - On obtient alors $\mathbf{y}(f, n) = \mathbf{C}(f) \mathbf{s}(f, n)$, où $\mathbf{C}(f) = \mathbf{B}(f) \mathbf{A}(f)$
 - $\mathbf{C}(f)$ est non mélangée



- Problème : indéterminations (permutations et coefficients multiplicatifs) dans les matrices $\mathbf{C}(f)$
 - $\forall k$, identifier les indices k_f tels que $\forall f, y_{k_f}(f, n) = c_{k_f, k} s_k(f, n)$
 - identifier les coefficients multiplicatifs $c_{k_f, k}$
- Infinité de solutions \Rightarrow besoin de contraindre le modèle :
 - Hypothèses sur le mélange
 - continuité des réponses en fréquence $a_{mk}(f)$ par rapport à f
 - \rightarrow modèle de formation de voie et modèle anéchoïque
 - Hypothèses sur les sources
 - similarité des dynamiques temporelles des $\sigma_k^2(f, n)$



- Modèle de **formation de voie** :
 - Hypothèses : ondes planes, champ lointain, pas de réverbération, antenne linéaire
 - Modèle : $a_{mk}(f) = e^{-2i\pi f \tau_{mk}}$ où $\tau_{mk} = \frac{d_m}{c} \sin(\theta_k)$
 - Paramètres : position d_m des capteurs et angles θ_k des sources
- Modèle **anéchoïque** :
 - Hypothèses : sources ponctuelles, pas de réverbération
 - Modèle : $a_{mk}(f) = \alpha_{mk} e^{-2i\pi f \tau_{mk}}$ où $\alpha_{mk} = \frac{1}{\sqrt{4\pi r_{mk}}}$ et $\tau_{mk} = \frac{r_{mk}}{c}$
 - Paramètres : distances r_{mk} entre les capteurs et les sources





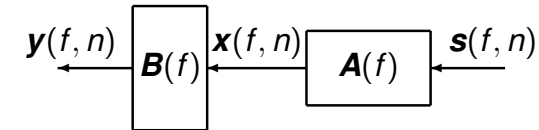
Partie VI

Mélanges sous-déterminés



Mélanges convolutifs sous-déterminés

- Cas courant en audio : mélanges mono ($K = 1$) ou stéréophoniques ($K = 2$), avec un nombre de sources $M > K$
- Modèle de mélange convolutif : $\mathbf{x}(f, n) = \mathbf{A}(f)\mathbf{s}(f, n)$ avec $M < K$
- Hypothèse : le modèle de mélange $\mathbf{A}(f)$ et le modèle de sources $\mathbf{R}_{ss}(f, n)$ sont connus
- Même dans ce cas, la séparation exacte est impossible, car il n'existe aucune matrice $\mathbf{B}(f)$ telle que $\mathbf{B}(f)\mathbf{A}(f) = \mathbf{I}_K$



Séparation par filtrage non stationnaire

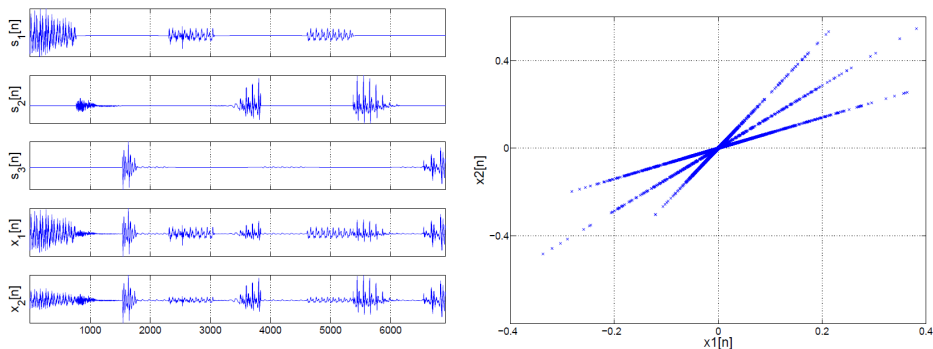
- On pose $\mathbf{y}(f, n) = \mathbf{B}(f, n)\mathbf{x}(f, n)$ où $\mathbf{B}(f, n) \in \mathbb{C}^{K \times M}$ dépend de n
- Estimateur *Minimum Mean Square Error* (MMSE) : on cherche $\mathbf{B}(f, n)$ qui minimise $\mathbb{E}[\|\mathbf{y}(f, n) - \mathbf{s}(f, n)\|^2]$
- Solution : $\mathbf{B}(f, n) = \mathbf{R}_{sx}(f, n)\mathbf{R}_{xx}(f, n)^{-1}$
où $\mathbf{R}_{xx}(f, n) = \mathbf{A}(f)\mathbf{R}_{ss}(f, n)\mathbf{A}(f)^H$ et $\mathbf{R}_{sx}(f, n) = \mathbf{R}_{ss}(f, n)\mathbf{A}(f)^H$
($(\cdot)^H$ désigne le conjugué hermitien)
- Remarque : $\mathbf{x}(f, n) = \mathbf{A}(f)\mathbf{y}(f, n)$ (reconstruction exacte)
- Cas particulier : mélanges monophoniques
 - Sans perte de généralité, on pose $\mathbf{A}(f) = [1, \dots, 1]$
 - On obtient $y_k(f, n) = \frac{\sigma_k^2(f, n)}{\sum_{l=1}^K \sigma_l^2(f, n)} x(f, n)$
 - \Rightarrow similaire au filtrage de Wiener



Méthode de séparation

1. Analyse TF des mélanges : $u_k(f, \cdot) = h_f * x_k$ et $x_k(f, n) = u_k(f, nT)$
2. Estimation de $\mathbf{A}(f)$ et $\sigma_k^2(f, n)$
 - modèle de mélange instantané
 - modèle de sources parcimonieuses
3. Calcul de $\mathbf{B}(f, n) = \mathbf{R}_{ss}(f, n)\mathbf{A}(f)^H (\mathbf{A}(f)\mathbf{R}_{ss}(f, n)\mathbf{A}(f)^H)^{-1}$
4. Estimation des signaux sources par $\mathbf{y}(f, n) = \mathbf{B}(f, n)\mathbf{x}(f, n)$
5. Synthèse TF des sources : $y_k(t) = \sum_{f=1}^F g_f(t - nT)y_k(f, n)$

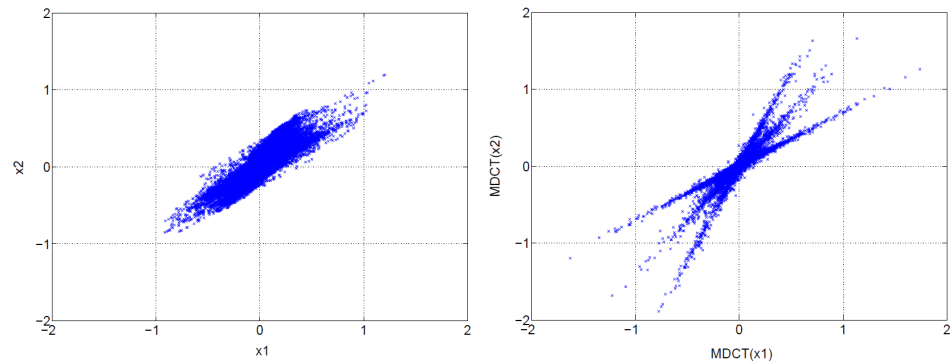
Cas d'un mélange stéréo linéaire instantané : $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$



(a) Signaux sources temporels et mélange stéréo correspondant

(b) Diagrammes de dispersion (x_1, x_2) en temps.

Cas d'un mélange stéréo linéaire instantané : $\mathbf{x}(f, n) = \mathbf{A}\mathbf{s}(f, n)$



(a) Échantillons temporels.

(b) Coefficients temps-fréquence après décomposition par MDCT.

- *Degenerate Unmixing Estimation Technique* (DUET)
- Modèle de mélange stéréo linéaire instantané : $\mathbf{x}(f, n) = \mathbf{A}\mathbf{s}(f, n)$
 - Sans perte de généralité, on suppose $\mathbf{A}_{(:,k)} = \begin{bmatrix} \cos(\theta_k) \\ \sin(\theta_k) \end{bmatrix} \forall k$
- Modèle de sources **parcimonieuses** :
 - $\forall f, n, \exists! k_{(f,n)}$ tel que $\sigma_{k_{(f,n)}}^2(f, n) > 0$, et $\forall l \neq k_{(f,n)}, \sigma_l^2(f, n) = 0$
- Si seule la source k est active en (f, n) , alors $\mathbf{x}(f, n) = \mathbf{a}_k s_k(f, n)$

1. Analyse TF des mélanges : $x_k(f, n) = (h_f * x_k)(nT)$
2. Estimation des paramètres θ_k et de la source active $k_{(f,n)}$
 - calcul de l'histogramme des angles des vecteurs $\mathbf{x}(f, n)$
 - détection des pics pour estimer les paramètres θ_k
 - détermination de la source active en (f, n) par proximité avec θ_k
3. Séparation des sources : pour tout k ,
 - estimation des images des sources par masquage binaire : $\mathbf{y}_k(f, n) = \mathbf{x}(f, n) \forall (f, n)$ tels que $k_{(f,n)} = k$ et $\mathbf{y}_k(f, n) = 0$ pour les autres points temps-fréquence (f, n)
 - estimation MMSE des sources : $y_k(f, n) = \hat{\mathbf{a}}_k(f)^\dagger \mathbf{y}_k(f, n)$
4. Synthèse TF des sources : $y_k(t) = \sum_{f=1}^F g_f(t - nT) y_k(f, n)$



Partie VII

Conclusion et perspectives



Conclusion

- Résumé
 - La séparation de sources nécessite de faire des hypothèses sur le mélange et sur les sources
 - Pour un mélange linéaire instantané (sur-)déterminé, l'hypothèse d'indépendance des sources est suffisante
 - Dans les autres cas, il faut modéliser le mélange et/ou les sources
- Perspectives
 - Mélanges non linéaires, non stationnaires (algorithmes adaptatifs)
 - Estimation du nombre de sources K
 - Séparation de sources informée

