

Maltraitement du signal

Partie 2 : le dopage

Gaël Mahé

Université Paris Cité

octobre 2022

Plan

- 1 Transformation d'histogramme
- 2 Parcimonisation jointe

Maltraitement du signal

Partie 2 : le dopage

1 Transformation d'histogramme

2 Parcimonisation jointe

Identification de systèmes non-linéaires

Systèmes audio non-linéaires :

- amplificateurs, haut-parleurs, microphones...
- NL liées à effets électriques, mécaniques et acoustiques
- Modélisation : modèles polynômiaux (sans mémoire) / filtres de Volterra (avec mémoire)

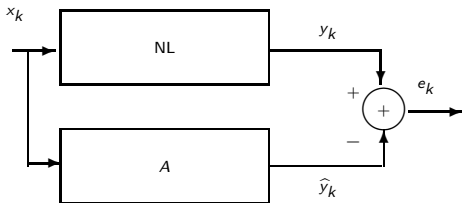
Classiquement, identification ou caractérisation *via*

- bruit blanc gaussien
- multitons ou sinusoides glissantes
- séquences à longueur maximale

Inconvénient : comportement du système dépend de l'entrée (*cf.* Klippel)

→ utiliser des signaux audio. Mais...

Identification d'un système sans mémoire (1)



- Modélisation polynômiale:

$$\hat{y}_k = A^T X_k, \text{ avec } X_k = (x_k, x_k^2, \dots, x_k^p)^T$$

- Le vecteur optimal A^{opt} minimise $E[(y_k - \hat{y}_k)^2]$:

$$\mathbf{C}_x A^{opt} = c_{Xy}, \text{ avec } \mathbf{C}_x = E[X_k X_k^T] \text{ et } c_{Xy} = E[X_k y_k]$$

- Méthode : algorithme adaptatif du type gradient stochastique, par ex. NLMS

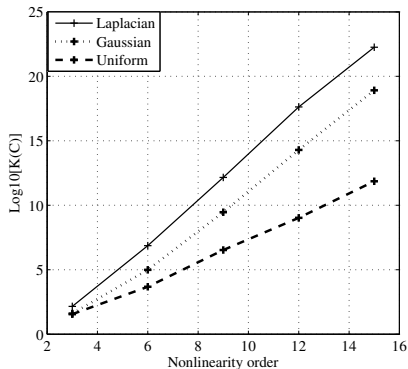
$$A_{k+1} = A_k + \frac{\mu}{\|X_k\|^2} e_k X_k, \quad (1)$$

Identification d'un système sans mémoire (2)

Performances d'identification liées au conditionnement de \mathbf{C}_x :

$$K(\mathbf{C}_x) = \log_{10} \left(\frac{|\lambda_{max}|}{|\lambda_{min}|} \right),$$

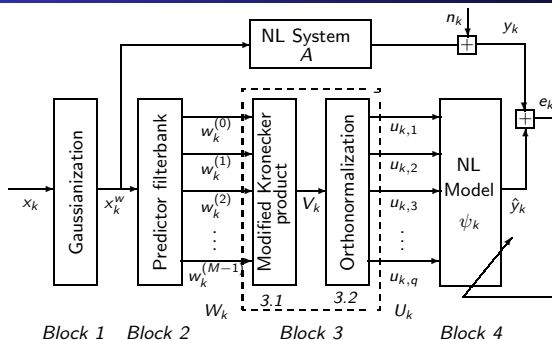
avec λ_{max} et λ_{min} valeurs propres max et min de \mathbf{C}_x .



Identification d'un système avec mémoire

- Soient N l'ordre polynômial, M la longueur de la mémoire.
- Modélisation par un filtre Volterra: $\hat{y}_k = A^T X_k$,
avec X_k constitué de tous les
 $x_k^{m_1} x_{k-1}^{m_2} \dots x_{k-M+1}^{m_M}$ / $m_1 + m_2 + \dots + m_M \leq N$
- Performances d'identification liées au conditionnement de $\mathbf{R}_x = E[X_k X_k^T]$:
 - croît exponentiellement avec N
 - borne sup : $K(\mathbf{R}_x) < [K(\mathbf{C}_x)]^M$

Comment améliorer le conditionnement des matrices ?



- U_k forme une base orthonormale
- $E[U_k U_k^T] =$ matrice unité (théoriquement) : conditionnement = 1

Comment orthonormaliser ?

- procédure de Gram-Schmidt valable quelle que soit la ddp
- utilisation des polynômes d'Hermite si le signal est gaussien

Conclusion : on aimerait que le signal soit plus gaussien.

Principes de la séparation de sources

- n sources $S_1 \dots S_n$
- matrice de mélange A de dimensions $p \times n$
- p mélanges $X_1 \dots X_p$

$$X = AS, \quad \text{avec } S = [S_1 \dots S_n]^T \text{ et } X = [X_1 \dots X_p]^T$$

Objectif : estimer S à partir de X sans connaître A

Difficulté dépend de

- $p \geq n$?
- propriétés connues des sources ?
(distribution, moments, bornes, modèle paramétrique...)
- mélange instantané ou convolutif ?
- mélange linéaire ou pas ?

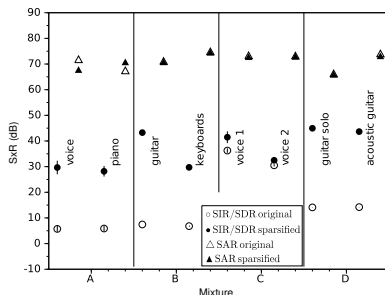
Applications : télécoms, biomédical, sismique, audio...

Séparation d'un mélange linéaire instantané déterminé

- **Théorème de Darmois** : Si les entrées sont indépendantes et au plus une d'entre elles est gaussienne, alors tout $Y = BX$ tel que les Y_i sont indépendants est égal à S à des permutations de sources et des facteurs d'échelle près.
- Algos de séparation par **maximisation d'indépendance** des Y_i
= independent component analysis (ICA)
- Performances dépendent de la distribution et de sa connaissance ou non.

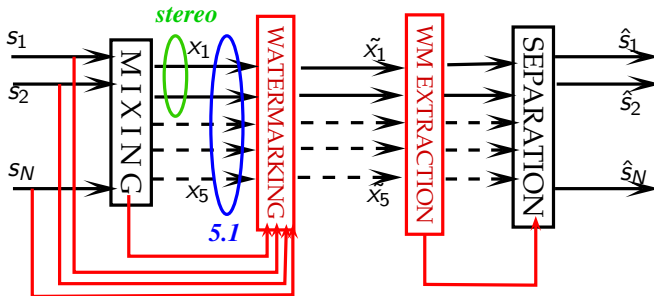
Séparation d'un mélange linéaire instantané déterminé (2)

En pratique, performances ICA faibles
pour des ddp intermédiaires entre gaussienne et laplacienne :



- Comment rendre les sources moins gaussiennes (plus supergaussiennes) ?
- Pour cela, les sources originales doivent être accessibles à un moment dans le processus
→ Contexte de la *séparation de sources informée*

Séparation de source informée

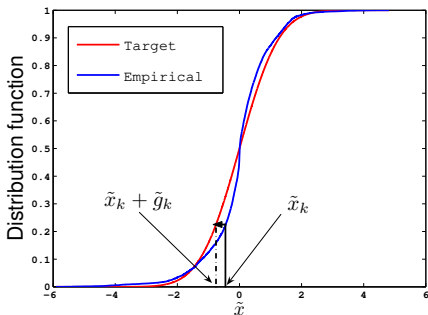


Égalisation d'histogramme

- Fonction de répartition empirique :

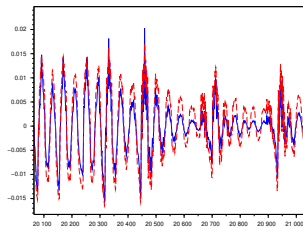
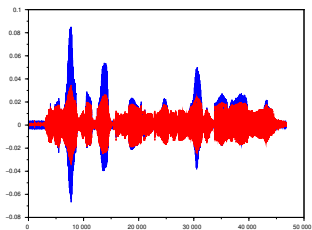
$$F_X^{emp}(x(n)) = \Pr[X \leq x(n)] = |\{X \leq x(n)\}|/N$$
- Fonction de répartition cible F_{target} , de mêmes moyenne et variance
- Egalisation d'histogramme comme en image :
remplacer chaque $x(n)$ par $z(n)$ tel que $z(n) = F_{target}^{-1}(F_X(x(n)))$
- *i.e.* ajouter un bruit g à chaque x , tel que

$$F_{target}(x(n) + g(n)) = F_X^{emp}(x(n))$$



Ex : gaussianisation

Gaussianisation de 3s de parole échantillonnée à 16kHz :



- Original
- Gaussianisé

Contrôle perceptif du reformage d'histogramme

- Histogramme et histogramme cumulatif d'une séquence $x(0) \dots x(N-1)$:

$$\forall k \in \mathbb{Z}, \begin{cases} f_x(k) = |\{x(n) \mid x(n) = k\}| \\ F_x(k) = \sum_{i=-\infty}^k f_x(i) = |\{x(n) \mid x(n) \leq k\}|, \end{cases} \quad (2)$$

- Soit f_{target} l'histogramme cible. Trouver une transfo $x \rightarrow z$ telle que :

$$\begin{cases} f_z \simeq f_{target} & (3) \\ w = z - x \text{ est inaudible} & (4) \end{cases}$$

- En d'autres termes,

$$\begin{cases} \min d(f_z, f_{target}) \text{ sous la contrainte :} & (5) \\ \gamma_w(m, \nu) < \gamma_{mask}(m, \nu) \quad \forall \text{trame } m, \text{ fréquence } \nu & (6) \end{cases}$$

Pourquoi est-ce difficile ?

- 1 Optimisation d'un histogramme défini sur tout le signal vs contraintes locales (un seuil de masquage par trame)
- 2 Optimisation exprimée à partir de la représentation temporelle vs contraintes exprimées dans le domaine fréquentiel

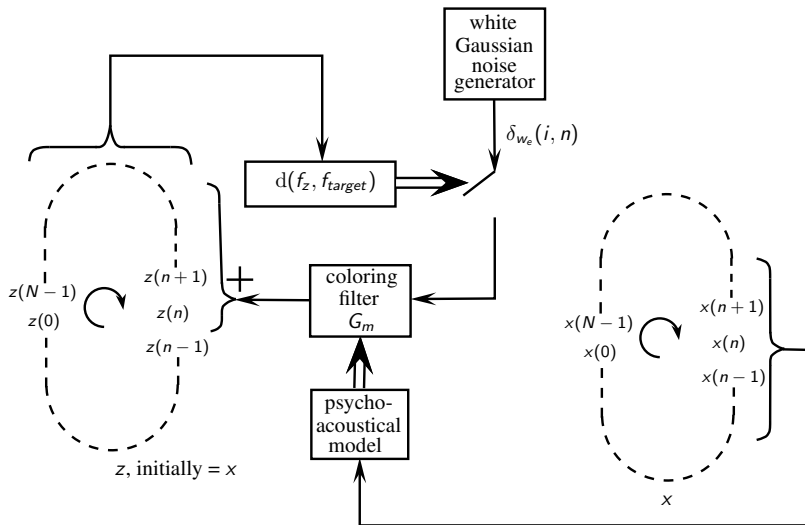
Idée

Traduction de la contrainte fréquentielle dans le domaine temporel :

$$w(n) = e(n) - \sum_{i=1}^p a_i w(n-i), \quad \text{avec } e \text{ bruit blanc}$$

...

Algo de reformage d'histogramme perceptivement contrôlé



Respect de la contrainte perceptive

- Pour q iterations, bruit de transformation $w = g_m * w_e$,
où $w_e(n) = \sum_{i=1}^q \delta(i, n) \delta w_e(i, n)$
où $\delta(i, n) = 0$ ou 1 selon decision d'ajouter $\delta w_e(i, n)$
- On peut montrer que w_e est un bruit blanc
- La contrainte $\gamma_w(\nu) < \gamma_{mask}(\nu)$ se traduit alors en :

$$|G_m(\nu)|^2 \sigma_{w_e} < \gamma_{mask}(\nu),$$

soit :

$$\begin{cases} |G_m(\nu)|^2 = \gamma_{mask}(\nu) \\ \sigma_{w_e} < 1 \quad \text{contrôlé par le choix de } q \text{ et } (\sigma_i)_{1 \leq i \leq q} \end{cases} \quad (7)$$

- L'algorithme s'arrête quand :
 f_z est suffisamment proche de f_{target} ou σ_{w_e} a atteint 1.

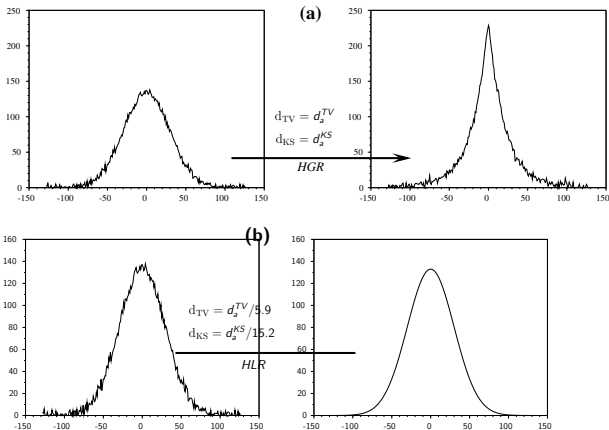
Choix d'implémentation

- H
- modèle psycho-acoustique
- q et $(\sigma_i)_{1 \leq i \leq q}$
- d

Choix de la distance entre histogrammes

Mesure de la dissimilarité entre 2 histogrammes f et g :

- Distance de la variation totale : $d_{TV}(f, g) = \frac{1}{2N} \sum_{k \in \mathbb{Z}} |f(k) - g(k)|$
- Dist. de Kolmogorov-Smirnov : $d_{KS}(f, g) = \frac{1}{N} \sup_{k \in \mathbb{Z}} |F(k) - G(k)|$



Application à la supergaussianisation (1)

Conditions expérimentales :

- Signaux mono-instrument de 10 à 15s, échantillonnage 32kHz
- Filtre G_m de longueur 50, approchant le seuil de masquage MPEG-1
- Histogramme cible : gaussienne généralisée avec facteur de forme divisé par 2

Application à la supergaussianisation (2)

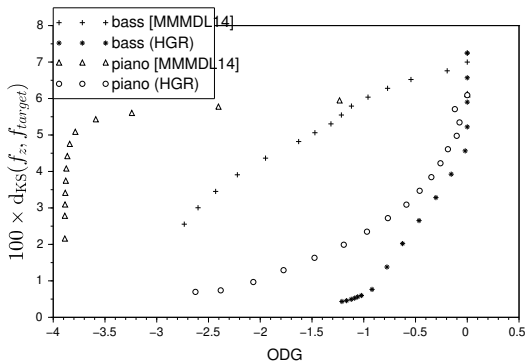
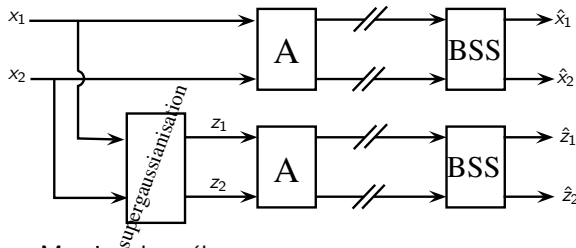


Figure: Distance de Kolmogorov-Smirnov à l'histogramme cible, $d_{KS}(f_z, f_{target})$, en fonction de l'ODG prédit par PEAQ.

En s'arrêtant dès que $\sigma_{w_e} > 1$,

- Basse, facteur de forme 1.6 \rightarrow 0.8, ODG final = -0.85
- Piano, facteur de forme 2.1 \rightarrow 1.05, ODG final = -1.1

L'application de l'application



- Matrice de mélange :

$$\mathbf{A} = \begin{pmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{pmatrix}$$

- BSS : séparation par l'algorithme FastICA

Résultats de supergaussianisation

Mixture	instruments	shape parameter	ODG mix
A	voice piano	1.8 \rightarrow 1.6 2.1 \rightarrow 1.6	-0.5
B	guitar keyboards	3 1.9 \rightarrow 1.3	-0.3
C	voice 1 voice 2	1 \rightarrow 0.9 1 \rightarrow 0.9	-0.6
D	guitar solo guitar acoustic	1.5 \rightarrow 1.5 1.4 \rightarrow 1.1	-0.8

Résultats de séparation

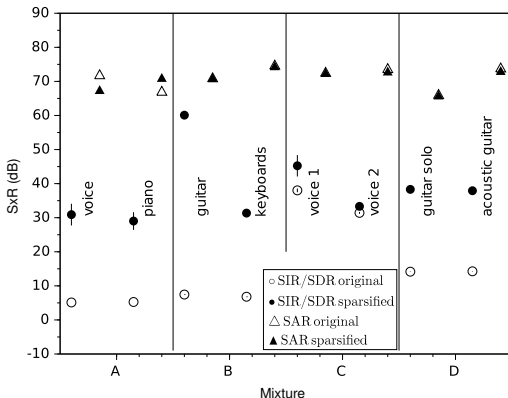


Figure: Performance of source separation with and without sparsification. Vertical bars represent confidence intervals (when applicable). In experiment B, the guitar signal was not sparsified. SIR = Source to Interference Ratio; SDR = Source to Distortion Ratio; SAR = Source to Artifact Ratio.

Maltraitement du signal

Partie 2 : le dopage

1 Transformation d'histogramme

2 Parcimonisation jointe

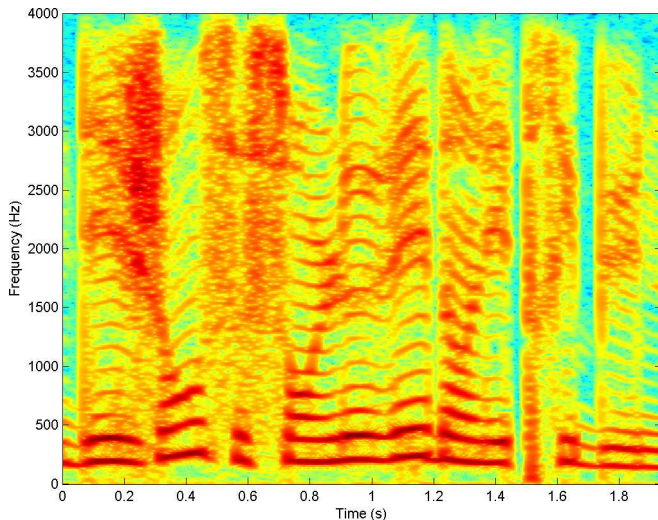
Séparation de sources sous-déterminée

- Nombre de capteurs $p <$ nombre de sources n
- Hypothèse de parcimonie : dans une certaine représentation, la plupart des coefficients d'une source sont nuls
- Hypothèse de parcimonie jointe : dans cette représentation, chaque coefficient du mélange est le mélange d'au plus p sources (ou peut être approché comme tel)
- \rightarrow techniques de sparse component analysis (SCA)
- Ex : SCA locale dans le domaine temps-fréquence (spectrogramme) :
 - Soit X le mélange, A la matrice de mélange
 - Pour un ensemble Λ de p sources, soit A_Λ la matrice A restreinte aux colonnes Λ
 - Estimation des p sources actives dans $X(f, t)$:

$$\hat{\Lambda} = \operatorname{argmax}_{|\Lambda|=p} \|A_\Lambda^{-1} X(f, t)\|_\tau, \quad \text{avec } 0 \leq \tau \leq 1$$

Pourquoi ça marche ?

Dans le domaine temps-fréquence,
les signaux audio sont naturellement parcimonieux :



Comment améliorer la parcimonie

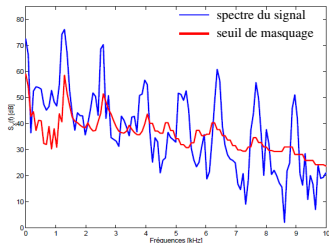


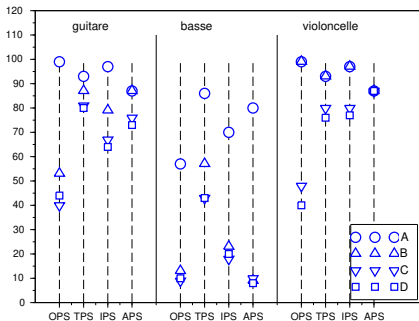
Figure: Seuil de masquage d'un son complexe.

- signal masque une partie de ses propres composantes → supprimer
- Voir J. Pinel and L. Girin. , "Sparsification" of audio signals using the MDCT/IntMDCT and a psychoacoustic model - application to informed audio source separation. In Proc. of the 42nd Audio Engineering Society Conference : Semantic Audio, Ilmenau, Germany, 2011.
→ mise à zéro de 75% du spectrogramme de chaque source sans distorsion audible, pour une fréquence d'échantillonnage de 44,1 kHz

La parcimonie ne suffit pas pour séparer les sources

Pour un mélange stéréo guitare-basse-violoncelle séparé par SCA, évaluation de la séparation *via* les Overall/Target/Interference/Artifact related Perceptual Scores estimés par PEASS.

- (C) Sources parcimonisées, information auxiliaire = matrice de mélange ;
- (D) Sources originales, information auxiliaire = matrice de mélange



Idée : modifier les sources pour forcer la SCA à réussir

- Mélange stéréo ($p = 2$)
- \mathcal{C} l'ensemble des paires d'entiers de $\llbracket 1; n \rrbracket$ ($n =$ nombre de sources)
- SCA estime les 2 sources actives dans $X(f, t)$:

$$\hat{\Lambda} = \operatorname{argmax}_{\mathcal{C}} \|\hat{S}^{(\Lambda)}\|_{\tau}$$

avec

$$\hat{S}^{(\Lambda)} = A_{\Lambda}^{-1} X = A_{\Lambda}^{-1} A S$$

- $\forall (f, t)$ où SCA échoue à trouver les sources dominantes $\Lambda \in \mathcal{C}$, on ajoute au vecteur des sources S un bruit ε
- estimation de $S + \varepsilon$ sous l'hyp. que les sources Λ sont dominantes :

$$\hat{S}_{\varepsilon}^{(\Lambda)} = A_{\Lambda}^{-1} A(S + \varepsilon)$$

- Trouver le ε tel que :

$$\begin{cases} \min \max \left(\frac{|(A\varepsilon)_1|}{M_1}, \frac{|(A\varepsilon)_2|}{M_2} \right) \\ \text{s.c. } \|\hat{S}_{\varepsilon}^{(\Lambda)}\|_{\tau} < \|\hat{S}_{\varepsilon}^{(\Lambda')}\|_{\tau} \quad \forall \Lambda' \in \mathcal{C} \setminus \{\Lambda\} \end{cases} \quad (8)$$

avec M_1 and M_2 seuils de masquage des canaux G/D

- Limite : pas toujours de solution et, si elle existe, le bruit ε peut être audible

Séparation de sources après forçage de la parcimonie jointe

Pour un mélange stéréo guitare-basse-violoncelle séparé par SCA, évaluation de la séparation *via* les Overall/Target/Interference/Artifact related Perceptual Scores estimés par PEASS.

- (B) Sources parcimonisées forcées
- (C) Sources parcimonisées
- (D) Sources originales

