

Résumé

On cherche à effectuer la séparation de mélanges instantanés de sources audio en utilisant une méthode basée sur l'algorithme de Décomposition Modale Empirique (ou EMD, pour Empirical Mode Decomposition). Cette approche nous permet, en particulier de traiter le cas sous déterminé (c'est à dire le cas où l'on a moins de capteurs que de sources).

Introduction

La séparation aveugle de sources est un problème qui consiste à retrouver des signaux "indépendants" à partir de leurs mélanges et cela sans connaissance *a priori* de la structure des mélanges ou des signaux sources.

Nous proposons ici une approche utilisant la décomposition modale empirique EMD (Empirical Mode Decomposition) des signaux observés [1, 2]. Elle s'applique aux signaux audio et aux signaux musicaux qui peuvent être modélisés par une somme de sinusoides amorties [7].

Formulation du problème

Nous pouvons formaliser le problème du mélange instantané comme suit:

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

$\mathbf{x}(t)$ représente le signal M -dimensionnel observé $\mathbf{x}(t) \stackrel{\text{def}}{=} [x_1(t), \dots, x_M(t)]^T$.

Ce signal $\mathbf{x}(t)$ correspond au mélange de N signaux sources $\mathbf{s}(t) \stackrel{\text{def}}{=} [s_1(t), \dots, s_N(t)]^T$ ($N > M$ pour le cas sous-déterminé) à travers une matrice "de mélange" $\mathbf{A} \stackrel{\text{def}}{=} [\mathbf{a}_1, \dots, \mathbf{a}_N]$.

Nous supposons que pour tout couple (i, j) les vecteurs \mathbf{a}_i et \mathbf{a}_j sont linéairement indépendants.

$\mathbf{w}(t)$ représente un bruit blanc additif.

Nous proposons une approche en deux étapes:

Analyse: décomposition EMD pour extraire les composantes modales.

Synthèse: classification et sélection pour regrouper les composantes d'une même source.

Analyse: décomposition EMD

Les signaux sources sont supposés décomposables en somme de sinusoides amorties $c_i^j(t)$, i.e:

$$s_i(t) = \sum_{j=1}^{l_i} c_i^j(t), \quad t = 0, \dots, T-1 \quad (2)$$

où les $c_i^j(t)$ sont des sinusoides amorties vérifiant:

$$\frac{\langle c_i^j | c_{i'}^{j'} \rangle}{\|c_i^j\| \|c_{i'}^{j'}\|} \approx 0 \quad \text{pour } (i, j) \neq (i', j') \quad (3)$$

où

$$\langle c_i^j | c_{i'}^{j'} \rangle \stackrel{\text{def}}{=} \sum_{t=0}^{T-1} c_i^j(t) c_{i'}^{j'}(t)^* \quad \text{et} \quad \|c_i^j\|^2 = \langle c_i^j | c_i^j \rangle \quad (4)$$

On obtient des estimées $\hat{c}_j^k(t)$ des composantes $c_j^k(t)$ en appliquant l'algorithme EMD sur les x_i qui s'écrivent:

$$x_i(t) = \sum_{j=1}^N a_{ij} s_j(t) = \sum_{j=1}^N \sum_{k=1}^{l_j} a_{ij} c_j^k(t) \quad (5)$$

Synthèse: Classification

Pour la synthèse des signaux sources on observe que:

$$\frac{\langle \mathbf{x} | c_i^j \rangle}{\|c_i^j\|^2} \stackrel{\text{def}}{=} \frac{1}{\|c_i^j\|^2} \begin{bmatrix} \langle x_1 | c_i^j \rangle \\ \vdots \\ \langle x_M | c_i^j \rangle \end{bmatrix} \approx \mathbf{a}_i \quad (6)$$

Nous pouvons associer chaque composante à une direction spatiale (vecteur colonne de \mathbf{A}) que l'on estime par:

$$\hat{\mathbf{a}}_i^j = \frac{\langle \mathbf{x} | \hat{c}_i^j \rangle}{\|\hat{c}_i^j\|^2} \quad (7)$$

Nous proposons de regrouper ces composantes par classification sur les vecteurs $\hat{\mathbf{a}}_i^k$ selon N classes.

On pourra reconstruire les sources initiales à une constante près en additionnant les différentes composantes d'une même classe.

Synthèse: Sélection

Nous remarquons, qu'en appliquant l'approche décrite précédemment sur toutes les sorties d'antenne $x_1(t), \dots, x_M(t)$, nous obtenons M estimées de chacune des sources.

Nous proposons une méthode "aveugle" de sélection pour choisir la meilleure des estimées en sélectionnant la plus énergétique, autrement dit:

$$\hat{s}_i(t) = \max_j \{ E_i^j = \sum_{t=0}^{T-1} \hat{s}_i^j(t)^2, \quad j = 1, \dots, M \} \quad (8)$$

où E_i^j représente l'énergie de la $i^{\text{ème}}$ source issue du $j^{\text{ème}}$ capteur.

On peut envisager d'autres méthodes de sélection basées sur la dispersion autour du centroïde de chaque classe, le nombre de composante pour chaque estimée, etc.

Simulation I

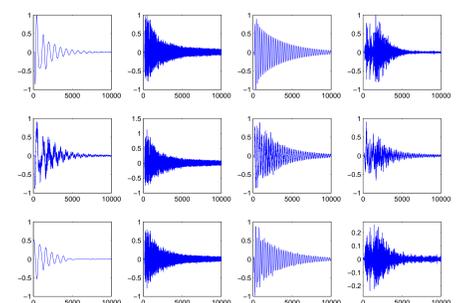


Figure 1: Exemple de séparation de signaux audio

La figure 1 représente les résultats obtenus en utilisant 4 sources audio échantillonnées à 44.1 khz et 3 capteurs. La première ligne représente les signaux sources originaux, la deuxième ligne illustre les signaux estimés par pseudo-inversion quand la matrice de mélange \mathbf{A} est connue, et la dernière donne les estimées des sources en utilisant notre algorithme.

Simulation II

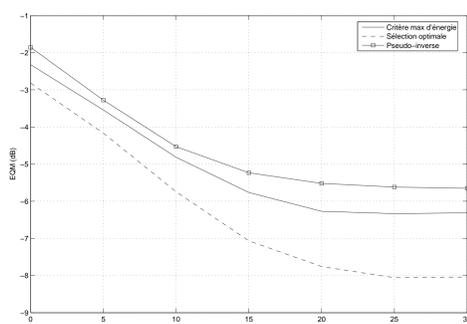


Figure 2: Performances de l'algorithme de séparation de sources audio

La figure 2 représente la variation de l'erreur quadratique moyenne (EQM) normalisée des estimées des signaux sources en fonction du rapport signal à bruit (RSB), pour un système composé de 4 sources ($N = 4$) échantillonnées à 44.1 khz et 3 capteurs ($M = 3$).

Conclusion

- Introduction d'une nouvelle méthode de séparation aveugle sous-déterminée de sources audio basée sur l'algorithme EMD.

- Présentation des différentes phases de cette méthode: **Analyse:** décomposition EMD et **Synthèse:** classification et sélection.

- Illustration des performances de la nouvelle méthode de séparation de sources.

Perspectives

- Amélioration des étapes de classification et de sélection, ainsi que l'utilisation d'autres méthodes de décomposition.

- Estimation du nombre N de sources.
- Méthode hybride: séparation par décomposition et filtrage spatial.

Références

- [1] N.E. Huang, Z. Shen, S.R. Long, M.L. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Chung and H.H. Liu, "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary times series analysis", *Proc. Roy. Soc. London A*, Vol. 454, pp. 903-995, 1998.
- [2] P. Flandrin, G. Rilling and P. Goncalvès, "Empirical mode decomposition as a filter bank", *IEEE Sig. Proc. Letters*, Vol. 11, N°2, pp. 112-114, 2004.
- [3] G. Govaert (sous la direction), "Analyse de données", *Lavoisier*, pp. 211-220, 2003.
- [4] A.K. Nandi editor, "Blind estimation using higher-order statistics", *Kluwer Academic Publishers*, Boston 1999.
- [5] L. Nguyen, A. Belouchrani, K. Abed-Meraim and B. Boashash, "Separating more sources than sensors using time-frequency distributions." in *Proc. ISSPA'01*, Vol. II, pp. 583-586, Kuala Lumpur, Malaysia, 2001.
- [6] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: demixing n sources from 2 mixtures," in *ICASSP'2000*, Vol. 5, pp. 2985-2988, Istanbul, Turkey, June 2000.
- [7] R. Boyer and K. Abed-Meraim, "Audio modeling based on delayed sinusoids." *Speech and Audio Processing, IEEE Transactions*, Vol. 12, N°2, pp. 110-120, Mars 2004.