



## TP Séparation de sources audio

**Roland Badeau**



Contexte académique } sans modifications

Voir page 3

M2 Mathématiques / Vision / Apprentissage - Analyse des signaux audiofréquences



Dans ce TP, vous allez réaliser une implémentation simple sous Matlab d'une variante de la méthode DUET (*Degenerate Unmixing Estimation Technique*), qui vise à séparer des sources sonores dans un mélange stéréophonique. On se place dans le cas d'un mélange linéaire instantané, sous-déterminé ( $M = 2$  capteurs et  $K > 2$  sources). La séparation est réalisée en exploitant l'information spatiale, et en supposant que les sources sont parcimonieuses dans le plan temps-fréquence (une seule source est active en chaque point temps-fréquence). La transformation utilisée sera la MDCT (*Modified Discrete Cosine Transform*), qui présente le double avantage d'être à valeurs réelles, et de produire une représentation plus parcimonieuse que la TFCT (Transformée de Fourier à Court Terme). Pour la calculer, vous utiliserez la boîte à outils *Linear Time/Frequency Toolbox* (`lftfat`), que vous téléchargerez sur le site du M2 MVA. Vous trouverez également sur ce site le fichier stéréo à traiter, nommé `mix.wav`. Pour utiliser les fonctions de la boîte à outils `lftfat`, vous devrez d'abord charger celle-ci à l'aide de la fonction `lftfatstart`.

## 1 Modèle de mélange et principe de la méthode DUET

La méthode DUET s'appuie sur le modèle de mélange suivant : en tout point temps-fréquence  $(f, n)$ ,

$$X(f, n) = S(f, n) A$$

où

- les vecteurs lignes  $X(f, n) = [X(f, n, 1), X(f, n, 2)]$  de dimension  $M = 2$  contiennent les MDCT des deux canaux stéréo  $x(t, m)$  du mélange observé ;
- la matrice  $A = [\cos(\theta), \sin(\theta)]$  de dimension  $K \times M$  est la matrice de mélange ;
- le vecteur colonne  $\theta$  de dimension  $K$  contient les angles  $\theta(k)$  des sources  $k$  ;
- les vecteurs lignes  $S(f, n) = [S(f, n, 1), \dots, S(f, n, K)]$  de dimension  $K$  contiennent les MDCT des  $K$  signaux sources inconnus.

Si seule la source  $k$  est active en  $(f, n)$ , le point d'affixe  $Z(f, n) = X(f, n, 1) + iX(f, n, 2) \in \mathbb{C}$  vérifie  $Z(f, n) = S(f, n, k) e^{i\theta(k)}$ , où  $S(f, n, k) \in \mathbb{R}$ . On remarque que son argument permet à la fois d'identifier la source  $k$  active en  $(f, n)$  et son angle  $\theta(k)$ . Le module de  $Z(f, n)$  permet quant à lui de déterminer la valeur de  $S(f, n, k)$ , au signe près. Pour lever l'indétermination de signe de  $S(f, n, k)$ , on supposera que  $\theta(k) \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ . Une fois la source  $k$  identifiée en tout point temps-fréquence, il suffit d'appliquer un masque binaire  $B(f, n, k)$  à  $X(f, n, m)$ , pour obtenir une estimation  $Y(f, n, m, k)$  de l'image stéréo de la source  $k$ . Le signal source est enfin reconstruit à l'aide de l'estimateur MMSE (*Minimum Mean Square Error*), qui vérifie  $S(f, n, k) = Y(f, n, 1, k) \cos(\theta(k)) + Y(f, n, 2, k) \sin(\theta(k))$ .

## 2 Travail demandé

1. Ouvrir le fichier `mix.wav` et le charger dans une matrice  $x(t, m)$ , de dimension  $T \times M$ , où  $M = 2$  et  $T$  est le nombre d'échantillons. Écouter le mélange avec un casque. Quel est le nombre  $K$  d'instruments que vous entendez ? Dans quelle direction les percevez-vous ?
2. Tracer le diagramme de dispersion temporelle, défini comme l'ensemble des points du plan de coordonnées  $(x(t, 1), x(t, 2))$  pour tous  $t$  (pour tracer un ensemble de points, on utilisera la fonction `plot` avec le paramètre `'x'`, et on normalisera les axes des abscisses et des ordonnées avec l'instruction `axis equal`). Parvenez-vous à distinguer les directions des sources ?
3. Calculer les MDCT  $X(f, n, m)$  des deux canaux stéréo  $x(t, m)$  (on utilisera la fonction `wmdct`, avec  $F = 512$  bandes fréquentielles, et la fenêtre `'sqrthann'`). Afficher les représentations temps-fréquences correspondantes,  $|X(f, n, m)|^2$  (on utilisera la fonction `plotwmdct`).



4. Tracer le diagramme de dispersion temps-fréquence, défini comme l'ensemble des points du plan d'affixe  $Z(f, n)$  pour tous  $f$  et  $n$ . Parvenez-vous à distinguer les directions des sources ? Comment l'expliquez-vous ?
5. Tracer l'histogramme des arguments des points d'affixe  $Z(f, n)$  pour tous  $f$  et  $n$  (on utilisera la fonction `atan` pour calculer les arguments modulo  $\pi$ , entre  $-\frac{\pi}{2}$  et  $+\frac{\pi}{2}$ , et la fonction `hist` pour calculer l'histogramme, dont on règlera le nombre de classes de façon à bien faire apparaître les directions des sources). En déduire une estimation des angles  $\theta(k)$  (on déterminera ces valeurs graphiquement à partir de l'histogramme).
6. Pour estimer la source active en tout point temps-fréquence  $(f, n)$ , on cherchera la source  $k$  dont l'angle  $\theta(k)$  est le plus proche de l'argument de  $Z(f, n)$ , modulo  $\pi$  (on pourra utiliser une mesure d'écart invariante modulo  $\pi$ , par exemple  $|\sin(\theta(k) - \angle Z(f, n))|$ ). Générer alors les masques binaires  $B \in \{0, 1\}$ , telles que  $B(f, n, k)$  vaut 1 si la source  $k$  est active en  $(f, n)$ , et 0 sinon.
7. Appliquer les masques  $B$  aux MDCT  $X(f, n, m)$  pour estimer les MDCT des images stéréo  $Y(f, n, m, k)$ . Reconstruire ensuite les images  $y(t, m, k)$  des signaux sources en appliquant la MDCT inverse (on utilisera la fonction `iwmdct`).
8. Écouter les  $K$  images stéréo reconstruites  $y(:, :, k)$ . Quels défauts percevez-vous ?
9. Calculer l'estimateur MMSE  $S(f, n, k)$  de la source  $k$ . Reconstruire les signaux sources  $s(t, k)$  en appliquant la MDCT inverse à  $S(f, n, k)$ . Écouter le résultat obtenu.
10. On souhaite à présent respatialiser les sources, c'est à dire resynthétiser le mélange  $x(t, m)$  en modifiant les angles  $\theta(k)$  (on remarquera qu'il est inutile de repasser dans le domaine MDCT). Essayez par exemple de permuter les directions des sources. Écoutez le résultat obtenu. Que peut-on dire concernant les défauts audibles ?





Contexte académique } sans modifications

***Par le téléchargement ou la consultation de ce document, l'utilisateur accepte la licence d'utilisation qui y est attachée, telle que détaillée dans les dispositions suivantes, et s'engage à la respecter intégralement.***

La licence confère à l'utilisateur un droit d'usage sur le document consulté ou téléchargé, totalement ou en partie, dans les conditions définies ci-après, et à l'exclusion de toute utilisation commerciale.

Le droit d'usage défini par la licence autorise un usage dans un cadre académique, par un utilisateur donnant des cours dans un établissement d'enseignement secondaire ou supérieur et à l'exclusion expresse des formations commerciales et notamment de formation continue. Ce droit comprend :

- le droit de reproduire tout ou partie du document sur support informatique ou papier,
- le droit de diffuser tout ou partie du document à destination des élèves ou étudiants.

Aucune modification du document dans son contenu, sa forme ou sa présentation n'est autorisée.

Les mentions relatives à la source du document et/ou à son auteur doivent être conservées dans leur intégralité.

Le droit d'usage défini par la licence est personnel et non exclusif. Tout autre usage que ceux prévus par la licence est soumis à autorisation préalable et expresse de l'auteur : [sitepedago@telecom-paristech.fr](mailto:sitepedago@telecom-paristech.fr)