ATIAM : Examen de Traitement du Signal Musical
Partie Applications et ouvertures

Geoffroy Peeters, Roland Badeau, Mardi 9 janvier 2024
RENDRE DEUX COPIES: l'une pour Geoffroy Peeters (exercices 1 à 3)
et l'autre pour Roland Badeau (exercices 4 et 5)

Les documents ne sont pas autorisés.

1 Cours "Cepstre"

Question 1 Expliquez ce qu'est le cepstre réel et son interprétation dans le cas d'un modèle source/filtre (on attend une réponse textuelle et avec des équations mathématiques).
Question 2 Que représentent les premiers coefficients du cepstre ? Que représente l'énergie aux basses céfrences ? aux hautes céfrences ?
Question 3 Comment peut-on utiliser le cepstre pour estimer la fréquence fondamentale f_0 d'un signal ?
Question 4 Quel est l'avantage de cette méthode d'estimation de f_0 par rapport à celle de l'auto-corrélation temporelle du signal ? Servez-vous de leurs expressions mathématiques dans le domaine fréquentiel.

2 Cours "Chromas"

Question 1 Expliquez ce que sont les Chromas.
Question 2 Pour le calcul des Chromas, la longueur de la fenêtre d'analyse est fonction de la fréquence $f_{min} \in \mathbb{R}^+$ de la note la plus basse considérée. Donner son expression mathématique en considérant une note MIDI notée $m_{min} \in \mathbb{N}^+$ (on attend une réponse textuelle et avec des équations mathématiques).
Question 3 Pourquoi dit-on que les Chromas sont sensibles au timbre des instruments de musique ?

3 Cours "Estimation multi-pitch"

Question 1 Qu'est-ce que la méthode du "produit spectral" pour estimer la fréquence fondamentale ? (on attend une réponse textuelle et avec des équations mathématiques)
Question 2 Quelles en sont ses principales hypothèses/limitations ?
Question 3 Expliquez le principe du lissage spectral ("spectral smoothing") utilisé dans la méthode d'estimation multi-pitch de [Klapuri, 2003, IEEE TASP].
Question 4 Comment le principe de "somme spectrale" est-il repris dans l'algorithme "Deep-Salience" de [Bittner et al., 2017, ISMIR] ?

4 Cours "NMF" : *regularized β-NMF*

In certain NMF applications, we are required to enforce sparsity constraints on the factor matrices W and H. This is often achieved by considering the following optimization problem:

$$(W^\star, H^\star) = \underset{W\geq 0, H\geq 0}{\arg\min} \Big[\sum_{f=1}^{F}\sum_{n=1}^{N} d_\beta(v_{fn}; \hat{v}_{fn}) + \lambda_W \sum_{f=1}^{F}\sum_{k=1}^{K} w_{fk} + \lambda_H \sum_{k=1}^{K}\sum_{n=1}^{N} h_{kn}\Big], \qquad (1)$$

where $\lambda_W > 0$, $\lambda_H > 0$, $\hat{v}_{fn} = \sum_{k=1}^{K} w_{fk}h_{kn}$ and $d_\beta(\cdot;\cdot)$ is the β-divergence that is defined as:

$$d_\beta(v; \hat{v}) = \frac{1}{\beta(\beta-1)}\Big(v^\beta + (\beta-1)\hat{v}^\beta - \beta v \hat{v}^{\beta-1}\Big). \qquad (2)$$

Question 1
What are the roles of $\lambda_W > 0$ and $\lambda_H > 0$ in this problem?
Question 2
Derive the multiplicative update rules for this particular problem.
Question 3
Let us assume that we obtain the optimal factors, $W^\star$ and $H^\star$. How can we use $W^\star$ and $H^\star$ for audio source separation? For music transcription?

5 Cours "Méthodes à haute résolution"

On rappelle que la méthode MUSIC consiste à diagonaliser la matrice de covariance $\boldsymbol{R}_{xx}$ du signal, et à déterminer les pôles $\{z_k\}_{k\in\{0\ldots K-1\}}$ en tant que solutions de l'équation

$$\|\boldsymbol{W}_\perp^H \boldsymbol{v}(z)\|^2 = 0 \qquad (3)$$

où $\boldsymbol{v}(z) = \big[1,\, z, \ldots,\, z^{n-1}\big]^T$, et la matrice $\boldsymbol{W}_\perp$, de dimension $n \times (n-K)$, contient les vecteurs propres de $\boldsymbol{R}_{xx}$ associés aux $n-K$ plus petites valeurs propres (engendrant ainsi l'espace bruit). Dans cet exercice, on suppose que tous les pôles du signal sont sur le cercle unité.
Question 1
Vérifier que l'équation (3) implique qu'ils sont également solutions de l'équation $P(z) = 0$, où

$$P(z) = z^{(n-1)}\, \boldsymbol{v}(1/z^*)^H \left(\boldsymbol{W}_\perp \boldsymbol{W}_\perp^H\right) \boldsymbol{v}(z).$$

Question 2
On définit la matrice $\boldsymbol{P} = \boldsymbol{W}_\perp \boldsymbol{W}_\perp^H$, et on remarquera que

$$P(z) = [z^{n-1},\, z^{n-2}, \ldots,\, z,\, 1]\, \boldsymbol{P} \begin{bmatrix} 1 \\ z \\ \vdots \\ z^{n-1} \end{bmatrix}.$$

Vérifier que la matrice $\boldsymbol{P}$ est à symétrie hermitienne et positive, et démontrer que $P(z)$ est un polynôme de degré au plus $2(n-1)$, dont les racines de module $\neq 1$ peuvent être regroupées par paires (si z est racine, $1/z^*$ l'est aussi).
Question 3
L'algorithme *root*-MUSIC consiste à calculer les racines de $P(z)$. D'après vous, comment pourrait-on en déduire les valeurs des pôles z_k ? (on prêtera attention au fait que le degré du polynôme P est supérieur à K).
Question 4
La méthode *spectral-MUSIC* vue en cours et testée en TP consiste à chercher les K maxima du pseudo-spectre $S(e^{i2\pi f}) = \frac{1}{\|\boldsymbol{W}_\perp^H \boldsymbol{v}(e^{i2\pi f})\|^2}$. Vérifier que les valeurs du pseudo-spectre pour les fréquences $f_k = \frac{k}{N_{\text{fft}}}$, où $k \in \{0, N_{\text{fft}} - 1\}$ et $N_{\text{fft}} \geq 2n-1$, peuvent être obtenues à l'aide de la TFD de longueur N_{fft} du signal constitué des coefficients du polynôme P, complétés par des zéros. Quel est l'inconvénient de cette approche par rapport à *root-MUSIC*?