Méthodes à haute résolution

M2 Mathématiques / Vision / Apprentissage
Analyse des signaux audiofréquences

Roland Badeau
TELECOM ParisTech

Modélisation sinusoïdale du signal audio

- Les sons qui engendrent une sensation de hauteur tonale bien définie ont une forme d'onde quasi-périodique
- Spectres d'harmoniques multiples du fondamental :
  - sons de parole voisés, produits par vibration quasi-périodique des cordes vocales
  - sons produits par des instruments à cordes ou à vent
- La propriété d'harmonicité n'est pas toujours vérifiée :
  - Certains instruments sont légèrement inharmoniques
  - Polyphonie : recouvrement des peignes d'harmoniques
  - Présence de paires ou de triplets de fréquences proches :
    - dissymétrie dans la géométrie d'une cloche
    - couplage entre cordes et chevalet dans une guitare
    - paires ou triplets de cordes dans un piano, et couplage des modes de vibration vertical et horizontal

Roland Badeau
TELECOM ParisTech

Partie I

Modèle paramétrique de signal

Roland Badeau
TELECOM ParisTech

Exponential Sinusoidal Model (ESM)

- Modulation d'amplitude exponentielle pour modéliser l'amortissement naturel des systèmes vibratoires libres
- Modèle réel : $s[t] = \sum_{k=0}^{K-1} a_k\, e^{\delta_k t} \cos(2\pi f_k t + \phi_k)$
- Modèle complexe : $s[t] = \sum_{k=0}^{K-1} a_k\, e^{\delta_k t}\, e^{i(2\pi f_k t + \phi_k)}$
- Écriture compacte : $s[t] = \sum_{k=0}^{K-1} \alpha_k\, {z_k}^t$ où
  - $\alpha_k = a_k e^{i\phi_k}$ est une *amplitude complexe*,
  - $z_k = e^{\delta_k + i2\pi f_k}$ est un *pôle complexe*.
- Hypothèses : pour tout $k \in \{0 \dots K-1\}$, $\alpha_k \neq 0$, $z_k \neq 0$, et tous les pôles z_k sont distincts deux à deux
- Le signal observé $x[t]$ est modélisé comme le signal $s[t]$ plus un bruit blanc gaussien complexe $b[t]$ de variance σ^2

Roland Badeau
TELECOM ParisTech

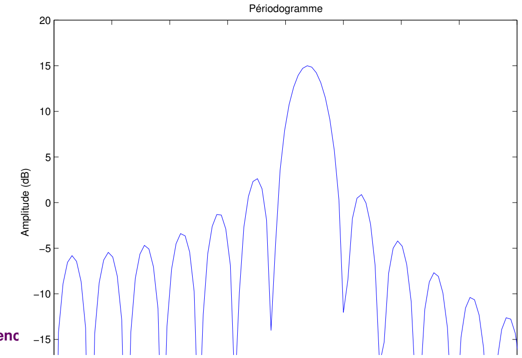
Estimation spectrale par analyse de Fourier

- Détection des pics dans la transformée de Fourier
- Avantages
  - existence d'un algorithme de calcul rapide (FFT)
  - méthode robuste d'estimation
- Inconvénients
  - résolution spectrale limitée par la longueur de la fenêtre
  - précision spectrale limitée par la taille de la transformée
  - compromis entre la largeur du lobe principal et la hauteur des lobes secondaires induites par la forme de la fenêtre
  - élargissement du pic en cas d'amortissement exponentiel

Problèmes de résolution

Signal de test :

- Fréquence d'échantillonnage : 8000 Hz
- Première sinusoïde : 440 Hz (la)
- Deuxième sinusoïde : 415,3 Hz (sol #)
- Pas d'amortissement, amplitudes égales à 1
- Longueur de la fenêtre rectangulaire : $N = 128$ (16 ms)
- Taille de la transformée : 1024 échantillons



Méthode du maximum de vraisemblance

- Principe général d'estimation paramétrique, asymptotiquement sans biais, consistant et efficace
- Il conduit à une estimation en trois étapes :
  - Estimation des pôles complexes : optimisation numérique d'une fonction de K variables complexes
  - Estimation des amplitudes complexes : par la méthode des moindres carrés
  - Estimation de la variance : puissance du résiduel
- Difficultés de la première étape :
  - complexité algorithmique
  - présence de nombreux maxima locaux
- Besoin de méthodes spécifiques pour les pôles complexes
- Les méthodes d'estimation paramétrique à haute résolution s'affranchissent des limites de l'analyse de Fourier

Partie II

Méthodes à haute résolution

Méthodes de prédiction linéaire

- Principe : tout signal satisfaisant la récurrence $s[t] - z_0\, s[t-1] = 0$ est de la forme $s[t] = \alpha_0\, {z_0}^t$
- Cas général : soit $P[z] \triangleq \prod_{k=0}^{K-1} (z - z_k) = \sum_{\tau=0}^{K} p_\tau\, z^{K-\tau}$.
- *Un signal discret $\{s[t]\}_{t\in\mathbb{Z}}$ satisfait l'équation de récurrence $\sum_{\tau=0}^{K} p_\tau\, s[t-\tau] = 0$ ssi il est de la forme $s[t] = \sum_{k=0}^{K-1} \alpha_k\, {z_k}^t$*
- Méthodes de Prony et de Pisarenko :
  - Estimer le polynôme $P[z]$ par prédiction linéaire
  - Extraire les racines de ce polynôme
- Inconvénient : performances médiocres en présence de bruit

TELECOM ParisTech

Représentation matricielle du signal

- Horizon d'observation : $t \in \{0 \dots N-1\}$, où $N > 2K$
- Matrice de données ($n > K$, $l > K$ et $N = n + l - 1$) :

$$\boldsymbol{S} = \begin{bmatrix} s[0] & s[1] & \dots & s[l-1] \\ s[1] & s[2] & \dots & s[l] \\ \vdots & \vdots & \vdots & \vdots \\ s[n-1] & s[n] & \dots & s[N-1] \end{bmatrix}$$

- Factorisation de la matrice $\boldsymbol{S}$: $\boldsymbol{S} = \boldsymbol{V}^n\, \boldsymbol{A}\, {\boldsymbol{V}^l}^T$, où
  - $\boldsymbol{V}^n$ est la matrice de Vandermonde de dimensions $n \times K$,

$$\boldsymbol{V}^n = \begin{bmatrix} 1 & 1 & \dots & 1 \\ z_0 & z_1 & \dots & z_{K-1} \\ {z_0}^2 & {z_1}^2 & \dots & {z_{K-1}}^2 \\ \vdots & \vdots & \vdots & \vdots \\ {z_0}^{n-1} & {z_1}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}$$

  - $\boldsymbol{V}^l$ est la matrice de Vandermonde de dimensions $l \times K$,
  - $\boldsymbol{A} = \mathrm{diag}(\alpha_0, \alpha_1, \dots, \alpha_{K-1})$ est une matrice diagonale de dimension $K \times K$.

TELECOM ParisTech

Matrice de corrélation

- On définit la matrice de corrélation $\boldsymbol{R}_{ss} = \frac{1}{l}\, \boldsymbol{S}\, \boldsymbol{S}^H$
- Alors $\boldsymbol{R}_{ss} = \boldsymbol{V}^n\, \boldsymbol{P}\, {\boldsymbol{V}^n}^H$, où $\boldsymbol{P} = \frac{1}{l}\, \boldsymbol{A}\, {\boldsymbol{V}^l}^T\, {\boldsymbol{V}^l}^*\, \boldsymbol{A}^H$
- La matrice $\boldsymbol{R}_{ss}$ est de rang K
- $\boldsymbol{R}_{ss}$ est diagonalisable en base orthonormée $\{\boldsymbol{w}_0 \dots \boldsymbol{w}_{n-1}\}$
- Ses valeurs propres $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{n-1} \geq 0$ satisfont
  - $\forall i \in \{0 \dots K-1\}$, $\lambda_i > 0$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda_i = 0$.
- On pose $\widehat{\boldsymbol{R}}_{bb} = \frac{1}{l}\boldsymbol{B}\, \boldsymbol{B}^H$ et $\boldsymbol{R}_{bb} = \mathbb{E}\left[\widehat{\boldsymbol{R}}_{bb}\right] = \sigma^2 \boldsymbol{I}_n$.
- De même, on pose $\widehat{\boldsymbol{R}}_{xx} = \frac{1}{l}\boldsymbol{X}\, \boldsymbol{X}^H$ et $\boldsymbol{R}_{xx} = \mathbb{E}\left[\widehat{\boldsymbol{R}}_{xx}\right]$.
- Alors $\boldsymbol{R}_{xx} = \boldsymbol{R}_{ss} + \sigma^2 \boldsymbol{I}_n$

TELECOM ParisTech

Espace signal et espace bruit

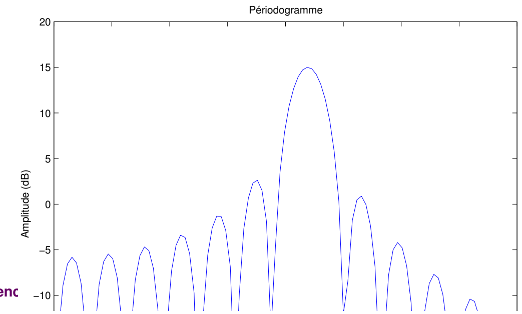
- Pour tout $i \in \{0 \dots n-1\}$, $\boldsymbol{w}_i$ est aussi vecteur propre de $\boldsymbol{R}_{xx}$ associé à la valeur propre $\lambda'_i = \lambda_i + \sigma^2$. Ainsi,
  - $\forall i \in \{0 \dots K-1\}$, $\lambda'_i > \sigma^2$;
  - $\forall i \in \{K \dots n-1\}$, $\lambda'_i = \sigma^2$.
- On note $\boldsymbol{W} = [\boldsymbol{w}_0 \dots \boldsymbol{w}_{K-1}]$, et $\boldsymbol{W}_\perp = [\boldsymbol{w}_K \dots \boldsymbol{w}_{n-1}]$
- Alors $\mathrm{Im}(\boldsymbol{W}) = \mathrm{Im}(\boldsymbol{V}^n)$ est appelé espace signal
- De même, $\mathrm{Im}(\boldsymbol{W}_\perp)$ est appelé espace bruit
- Les pôles $\{z_k\}_{k\in\{0\dots K-1\}}$ sont les solutions de l'équation $\left\|\boldsymbol{W}_\perp^H \boldsymbol{v}(z)\right\|^2 = 0$, où $\boldsymbol{v}(z) = [1, z, \dots, z^{n-1}]$
- La méthode MUSIC consiste à résoudre cette équation
- La méthode Spectral-MUSIC consiste à rechercher les K pics les plus élevés de la fonction $z \mapsto \frac{1}{\left\|\boldsymbol{W}_\perp^H \boldsymbol{v}(z)\right\|^2}$.

TELECOM ParisTech

Méthode *Spectral MUSIC*

Signal de test :

- Fréquence d'échantillonnage : 8000 Hz
- Première sinusoïde : 440 Hz (la)
- Deuxième sinusoïde : 415,3 Hz (sol #)
- Pas d'amortissement, amplitudes égales à 1
- Longueur de la fenêtre : $N = 128$ échantillons
- Valeurs des dimensions d'analyse : $n = 64$, $l = 65$, $K = 4$
- Taille du pseudo-spectre : 1024 échantillons



Algorithme ESPRIT

- Propriété d'invariance rotationnelle de $\boldsymbol{V}^n$:

$$\underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\substack{\boldsymbol{V}^n \\ n \times K}} \quad \underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\substack{\boldsymbol{V}^n{}_{\uparrow} \\ (n-1) \times K}}$$

$$\underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\boldsymbol{V}^n{}_{\uparrow}} \quad \underbrace{\begin{bmatrix} 1 & \dots & 1 \\ z_0 & \dots & z_{K-1} \\ \vdots & \dots & \vdots \\ {z_0}^{n-2} & \dots & {z_{K-1}}^{n-2} \\ {z_0}^{n-1} & \dots & {z_{K-1}}^{n-1} \end{bmatrix}}_{\boldsymbol{V}^n}$$

Algorithme ESPRIT

- Propriété d'invariance rotationnelle de $\boldsymbol{V}^n$: $\boldsymbol{V}^n{}_{\uparrow} = \boldsymbol{V}^n{}_{\downarrow}\,\boldsymbol{D}$
- Formule de changement de base : $\boldsymbol{V}^n = \boldsymbol{W}\,\boldsymbol{G}$
- Invariance rotationnelle de $\boldsymbol{W}(t)$: $\boldsymbol{W}(t)_{\uparrow} = \boldsymbol{W}(t)_{\downarrow}\,\boldsymbol{\Phi}(t)$ où $\boldsymbol{\Phi}(t) = \boldsymbol{G}\,\boldsymbol{D}\,\boldsymbol{G}^{-1}$ est appelée matrice spectrale
- Les valeurs propres de $\boldsymbol{\Phi}(t)$ sont les pôles $\{z_k\}_{k\in\{0\dots K-1\}}$
- La matrice $\boldsymbol{\Phi}$ vérifie $\boldsymbol{\Phi} = \left(\boldsymbol{W}_{\downarrow}^H \boldsymbol{W}_{\downarrow}\right)^{-1} \boldsymbol{W}_{\downarrow}^H \boldsymbol{W}_{\uparrow}$
- Algorithme ESPRIT :
  - calculer l'estimateur $\widehat{\boldsymbol{R}}_{xx}$ de la matrice $\boldsymbol{R}_{xx}$,
  - le diagonaliser et en déduire la matrice $\boldsymbol{W}$,
  - calculer $\boldsymbol{\Phi} = \left(\boldsymbol{W}_{\downarrow}^H \boldsymbol{W}_{\downarrow}\right)^{-1} \boldsymbol{W}_{\downarrow}^H \boldsymbol{W}_{\uparrow}$,
  - diagonaliser $\boldsymbol{\Phi}$ et en déduire les pôles $\{z_k\}_{k\in\{0\dots K-1\}}$.

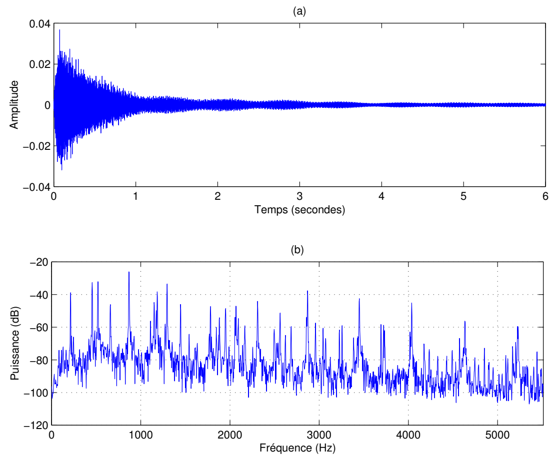
Estimation des amplitudes et des phases

- Soit $\boldsymbol{x}$ le vecteur $[x[0],\, x[1], \dots,\, x[N-1]]^T$ de dimension N
- On note $\boldsymbol{V}^N$ la matrice de Vandermonde contenant N lignes
- On note $\boldsymbol{\alpha} = [\alpha_0,\, \alpha_1, \dots,\, \alpha_{K-1}]^T$ le vecteur contenant les amplitudes complexes que l'on cherche à estimer
- Le principe du maximum de vraisemblance conduit à utiliser la méthode des moindres carrés : $\widehat{\alpha} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\| \boldsymbol{x} - \boldsymbol{V}^N \beta \right\|^2$
- La solution est $\widehat{\alpha} = \left(\boldsymbol{V}^{N^H} \boldsymbol{V}^N\right)^{-1} \boldsymbol{V}^{N^H} \boldsymbol{x}$
- On en déduit $\widehat{a}_k = |\widehat{\alpha}_k|$ et $\widehat{\phi}_k = \arg(\widehat{\alpha}_k)$

Partie III

Signaux à traiter dans le TP

Son de cloche



(a) Forme d'onde du signal
(b) Densité spectrale de puissance