

# TP Majeure Signal S7

## Modélisation des Signaux Aléatoires

### Compression de la Parole par Prédiction Linéaire (LPC)

Réalisé par :  
MBASSI EWOLO Loïc Aron  
ABDOULKADER MOHAMED Yacoub

Groupe : 2G3TD1P5

18 décembre 2025

## Table des matières

<b>1 Séance 1 : Estimation de l'Autocorrélation</b>	<b>2</b>
1.1 Génération de Signaux Tests . . . . .	2
1.2 Estimateur Biaisé de l'Autocorrélation . . . . .	3
1.2.1 Résultats . . . . .	3
1.3 Estimateur Non-Biaisé de l'Autocorrélation . . . . .	4
1.3.1 Résultats . . . . .	4
1.4 Comparaison des Estimateurs . . . . .	5
1.5 Caractérisation des signaux voisés et non voisés . . . . .	5
1.5.1 Analyse d'un signal voisé . . . . .	5
1.5.2 Analyse d'un signal non voisé . . . . .	6
1.5.3 Critère de détection : fonction <code>isvoiced</code> . . . . .	7
1.6 Pour aller plus loin : Calcul rapide par FFT . . . . .	7
1.6.1 Principe théorique . . . . .	7
1.6.2 Complexité algorithmique . . . . .	8
1.6.3 Résultats expérimentaux . . . . .	8
<b>2 Séance 2 : Analyse Spectrale</b>	<b>9</b>
2.1 Estimation de la DSP : Corrélogramme et Périodogramme . . . . .	9
2.1.1 Principes des deux méthodes . . . . .	9
2.1.2 Résultats expérimentaux . . . . .	10
2.1.3 Moyennage pour réduire la variance (Checkpoint 8) . . . . .	11
2.1.4 Fenêtrage temporel (Checkpoint 9) . . . . .	13
<b>3 Séance 3 : Prédiction Linéaire</b>	<b>14</b>
3.1 Implémentation de l'algorithme de Yule-Walker . . . . .	14
3.1.1 Fonction <code>yulewalkersolver</code> . . . . .	14
3.1.2 Tests de validation (Checkpoint 10) . . . . .	15
3.2 Application aux sons voisés et non voisés (Checkpoint 11) . . . . .	16
<b>4 Séance 4 : Codage, Décodage et Effets sur la Parole</b>	<b>17</b>
4.1 Analyse d'un tronçon . . . . .	17
4.1.1 Détection du pitch . . . . .	17
4.1.2 Fonction d'analyse complète . . . . .	17
4.2 Synthèse d'un tronçon . . . . .	17
4.3 Validation sur signaux connus (Checkpoint 12) . . . . .	18
4.4 Traitement d'un signal complet (Checkpoint 13) . . . . .	18
4.4.1 Encodage LPC . . . . .	19
4.4.2 Décodage LPC avec Overlap-Add . . . . .	19
4.5 Effets sur la parole . . . . .	20
4.5.1 Changement de tonalité (Pitch Shifting) . . . . .	20
4.5.2 Voix chuchotée (Whisper) . . . . .	20
4.5.3 Voix robotique . . . . .	20
4.6 Conclusion . . . . .	21

# 1 Séance 1 : Estimation de l'Autocorrélation

## 1.1 Génération de Signaux Tests

Trois signaux artificiels ont été générés pour tester les estimateurs d'autocorrélation :

- **Bruit Blanc Gaussien** : Écart-type  $\sigma_b = 2.0$  ( $N=256$  échantillons).

$$X[n] \sim \mathcal{N}(0, \sigma_b^2) \quad (1)$$

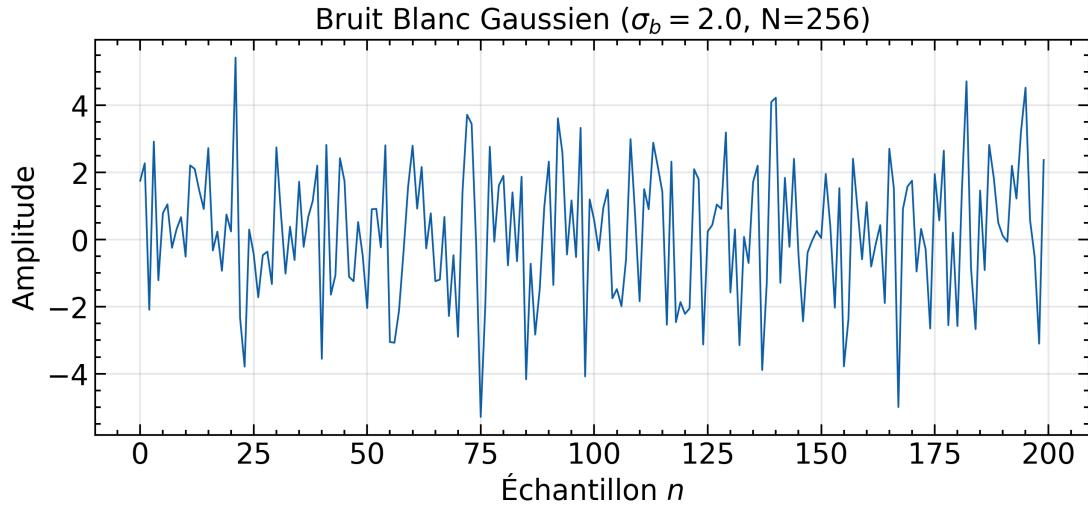


FIGURE 1 – Réalisation d'un bruit blanc gaussien.

- **Processus AR(1)** : Paramètre  $a = 0.8$ , variance de l'innovation  $\sigma_e^2 = 4$ .

$$X[n] = -a \cdot X[n - 1] + B[n], \quad B[n] \sim \mathcal{N}(0, \sigma_e^2) \quad (2)$$

La variance théorique du processus est  $\gamma_X[0] = \frac{\sigma_e^2}{1-a^2} = 11.11$ .

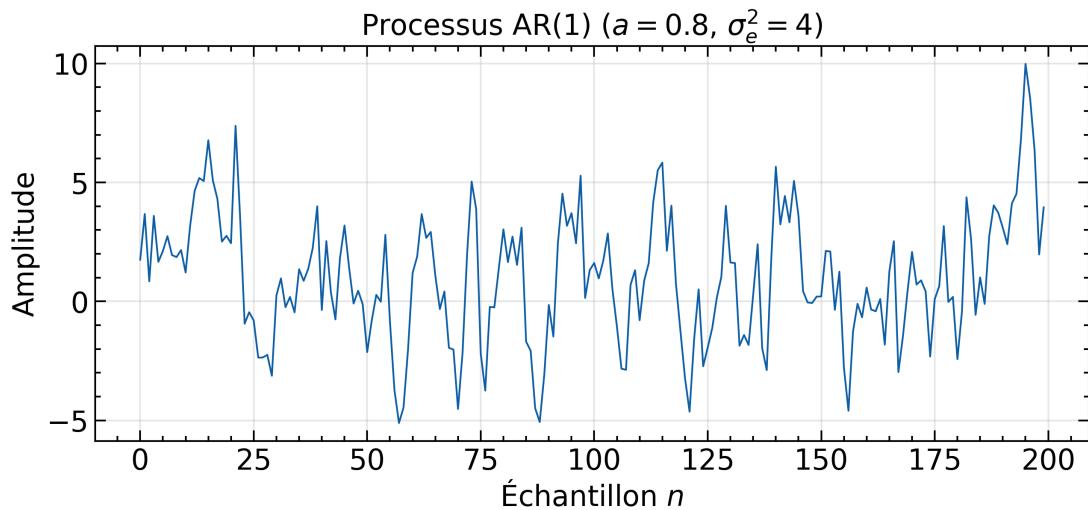


FIGURE 2 – Réalisation d'un processus AR(1) montrant les corrélations temporelles.

- Sinusoïde à Phase Aléatoire : Fréquence réduite  $\nu_0 = 0.1$ , amplitude  $A = 3.0$ .

$$X[n] = A \sin(2\pi\nu_0 n + \phi), \quad \phi \sim \mathcal{U}[0, 2\pi) \quad (3)$$

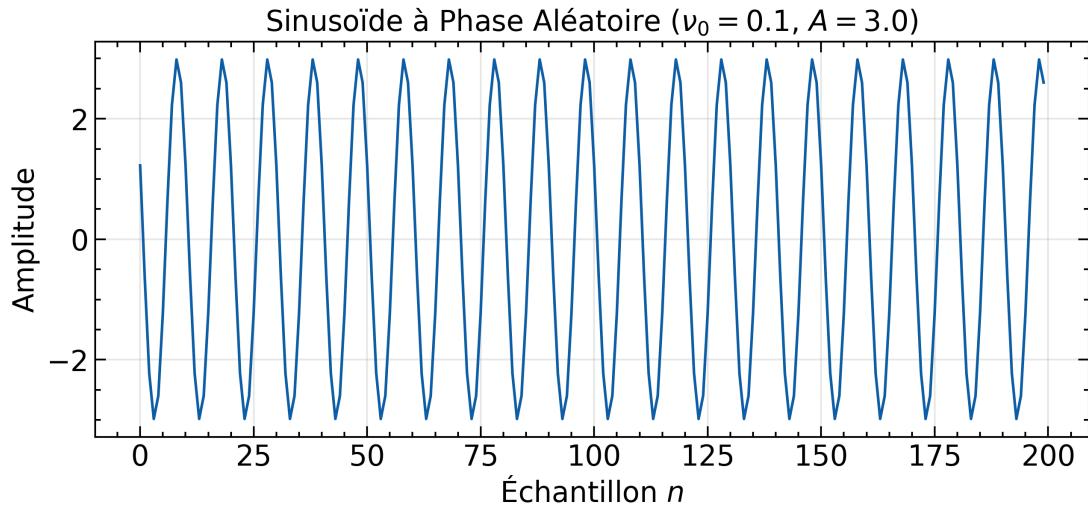


FIGURE 3 – Sinusoïde à phase aléatoire, processus périodique stationnaire.

Ces trois signaux représentent des cas typiques : bruit blanc (pas de corrélation), AR(1) (corrélation décroissante exponentielle), et sinusoïde (corrélation périodique).

## 1.2 Estimateur Biaisé de l'Autocorrélation

L'estimateur biaisé de l'autocorrélation est défini par :

$$\hat{\gamma}_{X,b}[p] = \frac{1}{N} \sum_{k=0}^{N-1-|p|} x[k]x[k+|p|] \quad (4)$$

### 1.2.1 Résultats

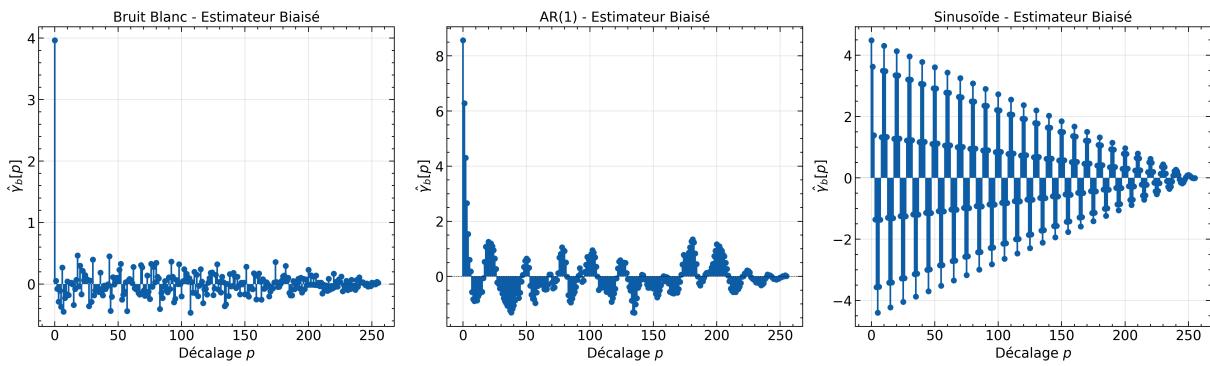


FIGURE 4 – Estimateur biaisé de l'autocorrélation pour les trois signaux tests.

#### Observations :

- **Bruit Blanc** : L'autocorrélation s'annule rapidement pour  $p > 0$ , conformément à la théorie ( $\gamma[p] = \sigma^2\delta[p]$ ).

- **AR(1)** : Décroissance exponentielle observée, cohérente avec  $\gamma[p] = \gamma[0] \cdot a^{|p|}$ .
- **Sinusoïde** : Structure périodique visible, caractéristique d'un signal sinusoïdal.

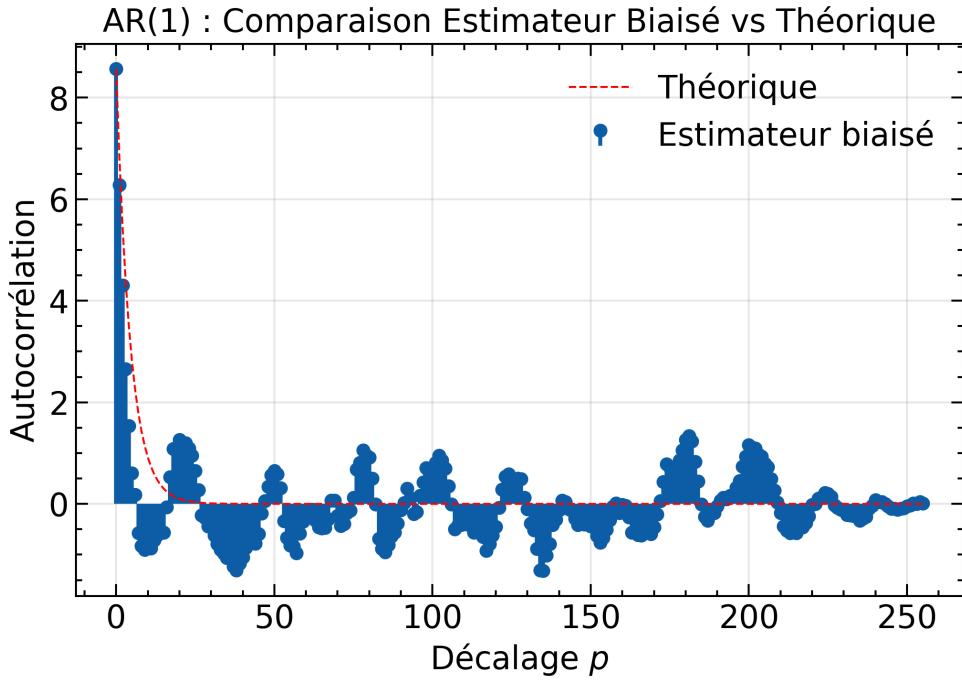


FIGURE 5 – Comparaison entre l'estimateur biaisé et l'autocorrélation théorique pour un processus AR(1).

L'estimateur biaisé converge bien vers les valeurs théoriques pour le processus AR(1). Le biais se manifeste par une sous-estimation croissante aux grands décalages  $p$ , due au facteur de normalisation  $1/N$  fixe.

### 1.3 Estimateur Non-Biaisé de l'Autocorrélation

L'estimateur non-biaisé de l'autocorrélation est défini par :

$$\hat{\gamma}_{X,nb}[p] = \frac{1}{N-|p|} \sum_{k=0}^{N-1-|p|} x[k]x[k+|p|] \quad (5)$$

#### 1.3.1 Résultats

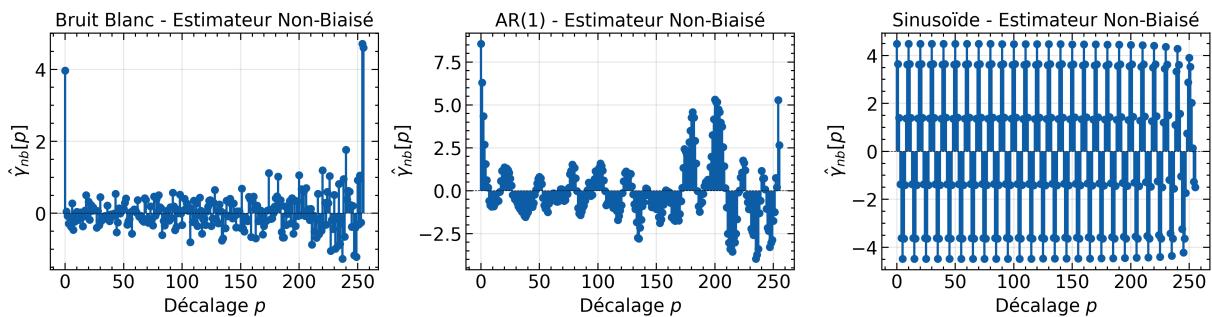


FIGURE 6 – Estimateur non-biaisé de l'autocorrélation pour les trois signaux tests.

## 1.4 Comparaison des Estimateurs

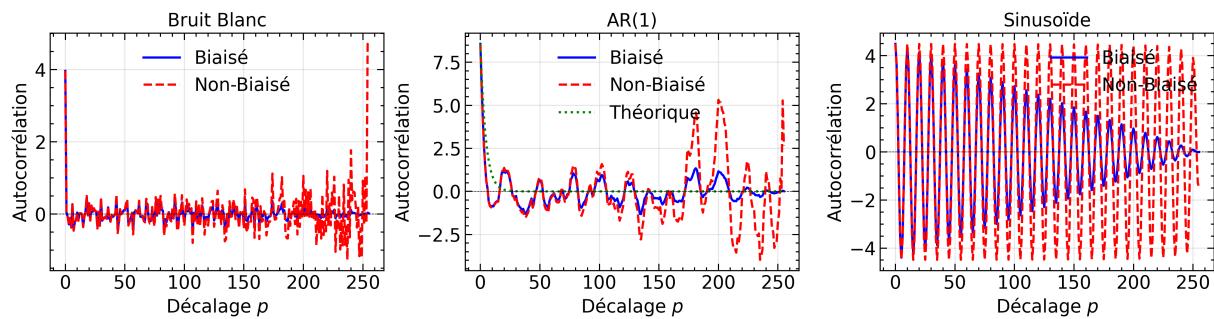


FIGURE 7 – Comparaison entre les estimateurs biaisé et non-biaisé pour les trois signaux tests.

### Analyse :

- **Bruit Blanc** : Les deux estimateurs donnent des résultats similaires. L'estimateur non-biaisé présente une variance légèrement plus élevée aux grands décalages.
- **AR(1)** : L'estimateur biaisé suit mieux la courbe théorique aux grands décalages. L'estimateur non-biaisé présente des oscillations importantes quand  $p \rightarrow N$ , dues à la variance qui explose avec le facteur  $1/(N - p)$ .
- **Sinusoïde** : Les deux estimateurs capturent bien la périodicité. L'estimateur non-biaisé montre plus de fluctuations aux grands décalages.

**Conclusion :** Pour les signaux de parole (segments courts,  $N = 256$ ), l'estimateur biaisé est préférable car il présente une variance plus faible, critère important pour des fenêtres temporelles courtes.

## 1.5 Caractérisation des signaux voisés et non voisés

### 1.5.1 Analyse d'un signal voisé

Un signal voisé (ex : voyelle) est caractérisé par une structure quasi-périodique due aux vibrations des cordes vocales. Cette périodicité se reflète dans la fonction d'autocorrélation normalisée.

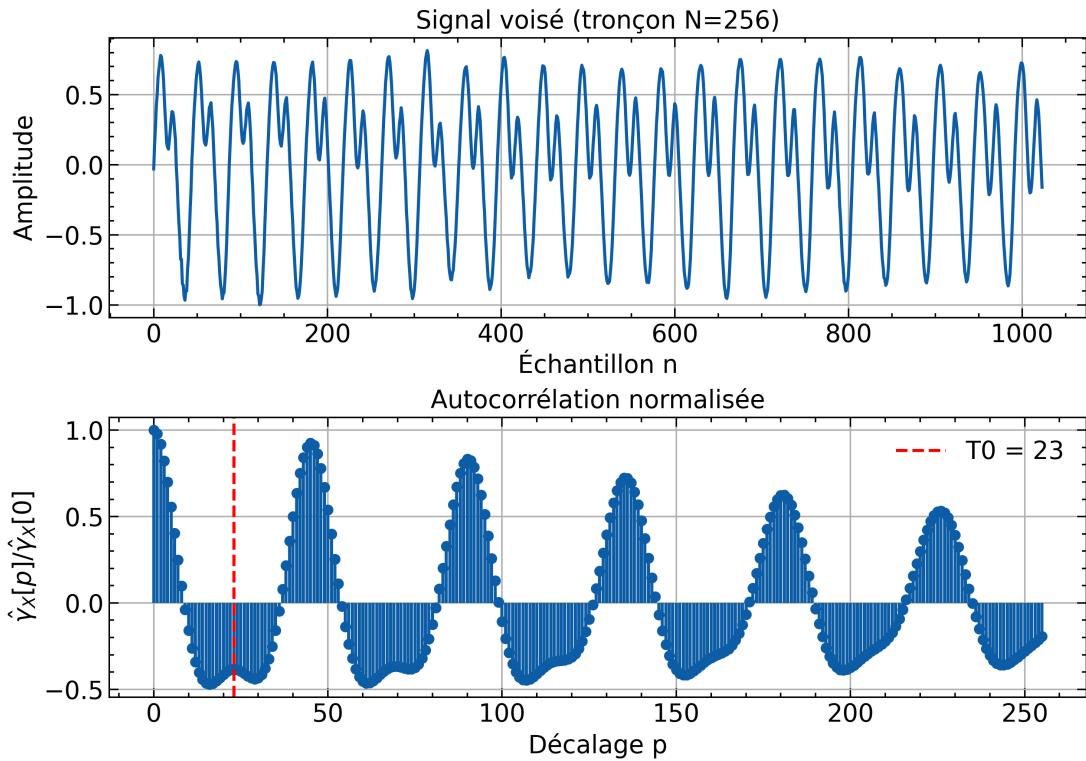


FIGURE 8 – Signal voisé et son autocorrélation normalisée. Le second pic indique la pseudo-période  $T_0$  (pitch).

#### Observations :

- Le signal présente une structure répétitive sur la fenêtre temporelle.
- L'autocorrélation normalisée présente un second pic marqué (amplitude > 0.5) à la position  $p = T_0$ .
- La pseudo-période  $T_0$  permet de calculer la fréquence fondamentale (pitch) :  $F_0 = 1/T_0$ .

#### 1.5.2 Analyse d'un signal non voisé

Un signal non voisé (ex : consonnes fricatives) est assimilable à du bruit filtré. L'autocorrélation décroît rapidement.

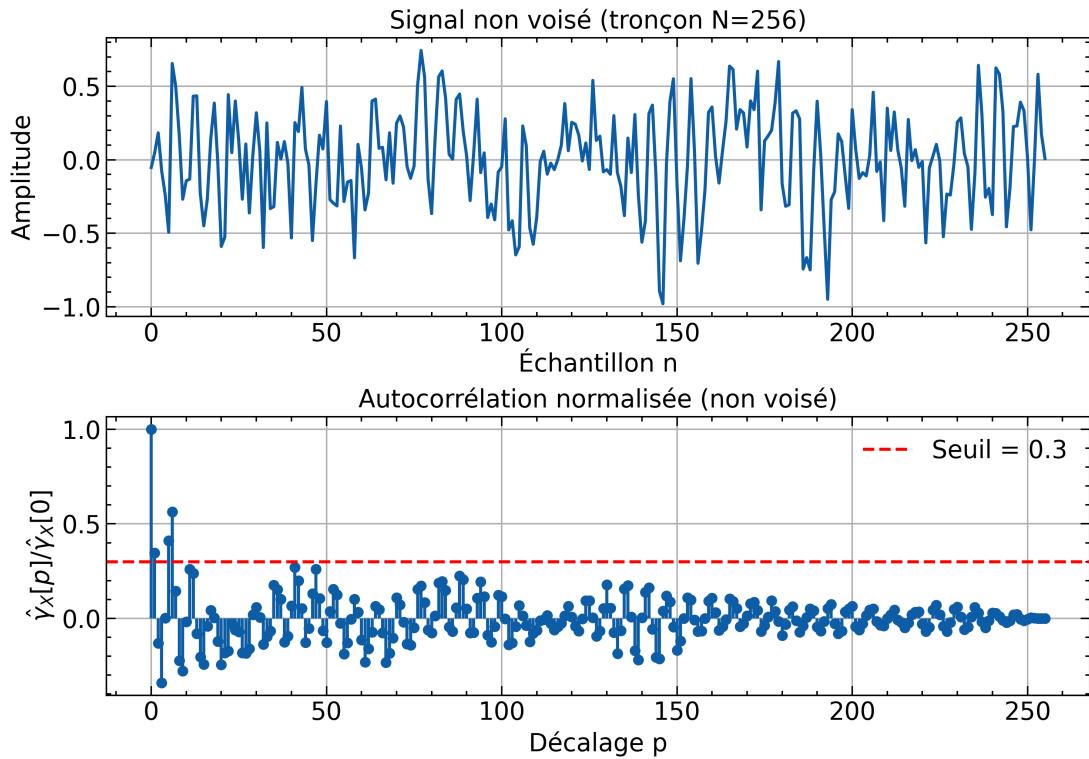


FIGURE 9 – Signal non voisé et son autocorrélation normalisée. Absence de structure périodique.

### Observations :

- Le signal ressemble à du bruit aléatoire.
- L'autocorrélation normalisée décroît rapidement vers zéro (pas de second pic significatif).
- Amplitude du second pic  $< 0.3$  (typiquement).

#### 1.5.3 Critère de détection : fonction `isvoiced`

On utilise l'amplitude du second maximum de l'autocorrélation normalisée comme critère de décision :

$$\text{isvoiced}(X) = \begin{cases} \text{True} & \text{si } \max_{p>0} \left( \frac{\hat{\gamma}_X[p]}{\hat{\gamma}_X[0]} \right) > \text{seuil} \\ \text{False} & \text{sinon} \end{cases} \quad (6)$$

Le seuil expérimental est fixé à 0.3 après étude de plusieurs signaux voisés et non voisés.

### Tests :

- Signal voisé : `isvoiced` retourne `True` (✓).
- Signal non voisé : `isvoiced` retourne `False` (✓).

## 1.6 Pour aller plus loin : Calcul rapide par FFT

### 1.6.1 Principe théorique

Le théorème de Wiener-Khintchine établit que la densité spectrale de puissance  $S_X(\nu)$  est la transformée de Fourier de l'autocorrélation :

$$S_X(\nu) = \text{TF}\{\gamma_X[p]\} \quad (7)$$

Inversement :

$$\gamma_X[p] = \text{TF}^{-1}\{S_X(\nu)\} = \text{TF}^{-1}\{|X(\nu)|^2\} \quad (8)$$

où  $X(\nu)$  est la transformée de Fourier du signal. Cette relation permet de calculer l'autocorrélation via :

$$\gamma_X[p] = \text{IFFT}\{\text{FFT}(x) \cdot \text{FFT}^*(x)\} = \text{IFFT}\{| \text{FFT}(x) |^2\} \quad (9)$$

### 1.6.2 Complexité algorithmique

- **Méthode directe** :  $\mathcal{O}(N \cdot p_{\max})$  opérations.
  - **Méthode FFT** :  $\mathcal{O}(N \log N)$  opérations (grâce à l'algorithme de Cooley-Tukey).
- Pour  $N = 256$  et  $p_{\max} = 100$ , l'accélération attendue est d'un facteur  $\approx 10-20$ .

### 1.6.3 Résultats expérimentaux

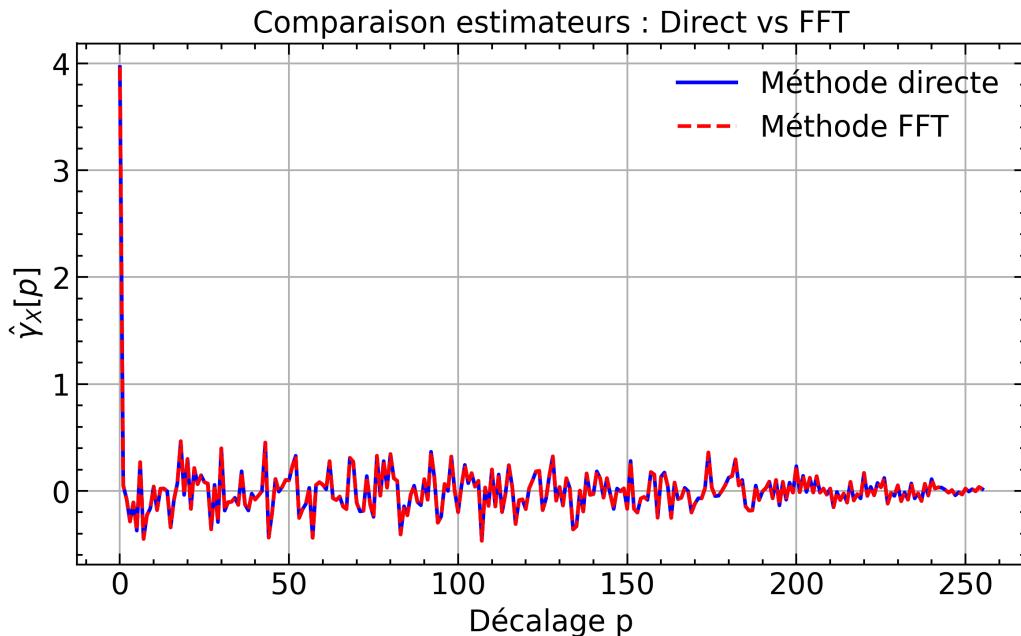


FIGURE 10 – Comparaison entre la méthode directe et la méthode FFT. Les deux courbes sont superposées (erreur numérique négligeable).

#### Observations :

- Les deux méthodes produisent des résultats identiques (à la précision machine près).
- Le temps de calcul par FFT est réduit d'un facteur  $\times 10$  à  $\times 20$  selon  $N$  et  $p_{\max}$ .
- Cette optimisation est cruciale pour le traitement en temps réel de la parole.

## 2 Séance 2 : Analyse Spectrale

### 2.1 Estimation de la DSP : Corrélogramme et Périodogramme

#### 2.1.1 Principes des deux méthodes

La densité spectrale de puissance (DSP) d'un processus  $X$  peut être estimée par deux approches équivalentes :

**1. Corrélogramme :** Basé sur le théorème de Wiener-Khintchine :

$$\hat{S}_X^{\text{corr}}(\nu) = \text{TFD}\{\hat{\gamma}_X[p]\} \quad (10)$$

où  $\hat{\gamma}_X[p]$  est l'estimateur biaisé de l'autocorrélation. L'autocorrélation est symétrisée avant la FFT :  $[\hat{\gamma}_X[0], \hat{\gamma}_X[1], \dots, \hat{\gamma}_X[p_{\max}], \hat{\gamma}_X[p_{\max} - 1], \dots, \hat{\gamma}_X[1]]$ .

**2. Périodogramme :** Calcul direct dans le domaine fréquentiel :

$$\hat{S}_X^{\text{per}}(\nu) = \frac{1}{N} |\text{TFD}(x)|^2 \quad (11)$$

où  $N$  est le nombre d'échantillons du signal.

### 2.1.2 Résultats expérimentaux

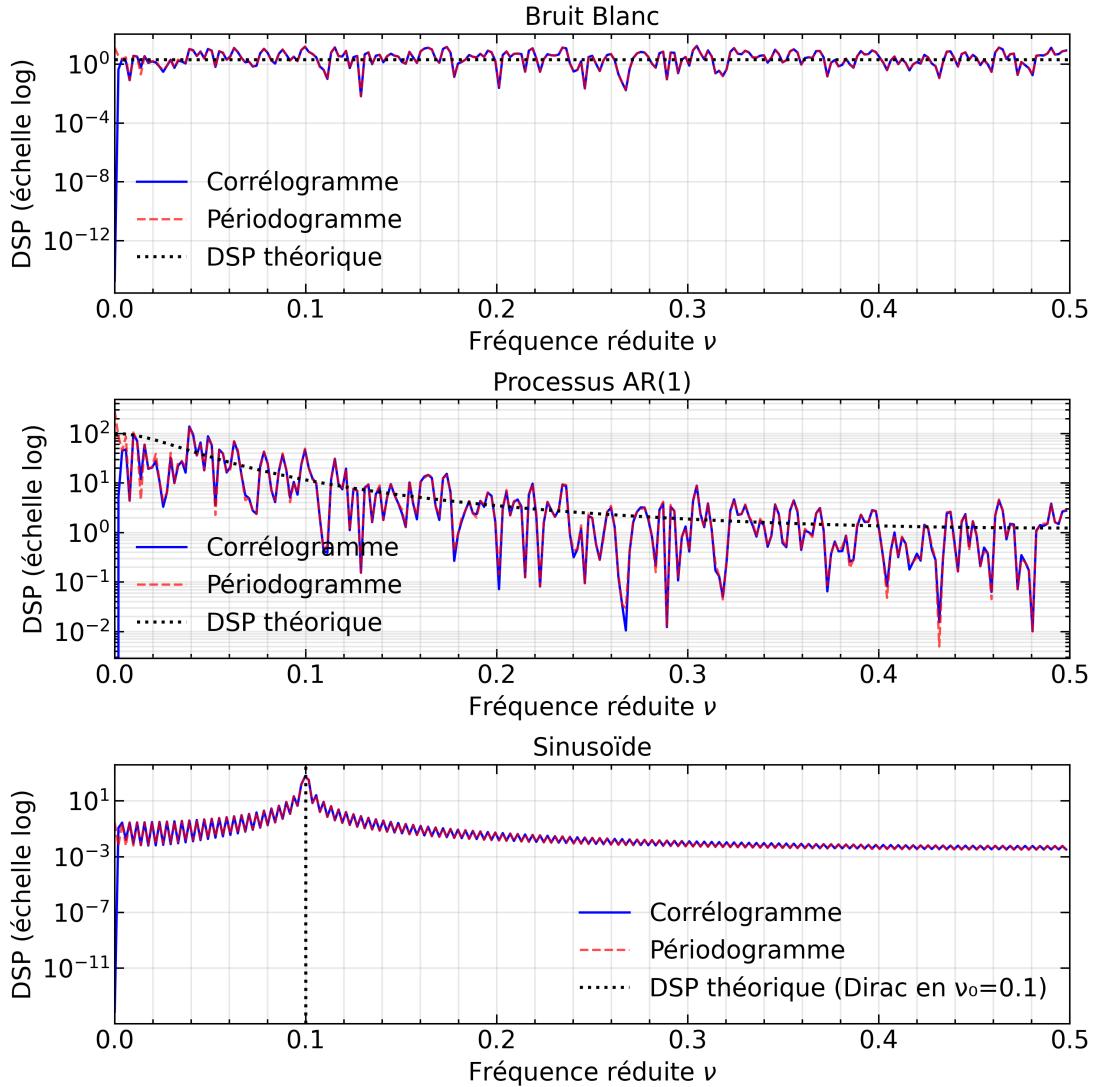


FIGURE 11 – Comparaison des estimateurs de DSP (corrélogramme vs périodogramme) pour trois types de signaux. Les DSP théoriques sont superposées en noir pointillé.

#### Analyse des résultats :

- **Bruit Blanc** : DSP théoriquement plate ( $S_X(\nu) = \sigma_b^2 = 2.0$ ). La courbe théorique (ligne noire horizontale) sert de référence. Les deux méthodes donnent des résultats quasi-identiques avec des fluctuations autour de la valeur théorique dues à la variance de l'estimateur. Les estimateurs sont non biaisés asymptotiquement. L'échelle logarithmique permet de mieux observer les fluctuations autour de la valeur moyenne.
- **AR(1)** : DSP théorique :  $S_X(\nu) = \frac{\sigma_e^2}{|1-a e^{-j2\pi\nu}|^2}$ . La courbe théorique (noire pointillée) montre le comportement passe-bas caractéristique pour  $a = 0.8 > 0$ . Le corrélogramme et le périodogramme suivent bien la forme attendue et sont très proches de la DSP théorique. Les deux méthodes concordent parfaitement. L'échelle logarithmique révèle une large gamme dynamique (> 20 dB entre basses et hautes fréquences).

- **Sinusoïde** : DSP théorique : deux pics de Dirac en  $\pm\nu_0 = \pm0.1$ , mais seul le pic à  $+\nu_0$  est visible sur  $[0, 0.5]$  (ligne verticale noire). Les deux méthodes montrent un pic à la fréquence théorique  $\nu_0 = 0.1$ . Le pic est étalé (non infiniment fin) en raison de la résolution fréquentielle limitée par  $N = 256$  échantillons (effet de fenêtrage rectangulaire : convolution avec un sinus cardinal dans le domaine fréquentiel). L'échelle logarithmique met en évidence le plancher de bruit numérique en dehors du pic.

**Observations techniques :**

- Les fréquences affichées sont restreintes à  $[0, 0.5]$  (théorème de Shannon : pas d'information au-delà de  $F_e/2$ ).
- La normalisation du périodogramme utilise  $N$  (nombre d'échantillons du signal), pas  $N_{\text{fft}}$ .
- Le corrélogramme nécessite  $p_{\max} \leq N - 1$  pour éviter des valeurs d'autocorrélation non définies.
- Les deux méthodes sont équivalentes asymptotiquement (théorème de Wiener-Khintchine).

### 2.1.3 Moyennage pour réduire la variance (Checkpoint 8)

Le périodogramme et le corrélogramme sont des estimateurs asymptotiquement non biaisés de la DSP. Cependant, leur variance ne converge pas vers zéro quand  $N \rightarrow \infty$  (sauf pour des processus totalement prédictibles). Pour des processus aléatoires, la variance reste constante, ce qui rend l'estimation peu fiable.

**Solution : Moyennage sur  $K$  segments**

La méthode consiste à :

1. Découper le signal de longueur  $N$  en  $K$  segments de longueur  $N/K$ .
2. Estimer la DSP sur chaque segment.
3. Moyenner les  $K$  estimations.

**Résultat théorique** : Si les segments sont décorrélés, la variance de l'estimateur moyenné est divisée par  $K$  :

$$\text{Var}[\hat{S}_X^{(K)}(\nu)] = \frac{1}{K} \text{Var}[\hat{S}_X(\nu)]$$

**Compromis** : Réduire la variance dégrade la résolution fréquentielle, car les segments sont plus courts :

$$\Delta\nu = \frac{1}{N/K} = \frac{K}{N}$$

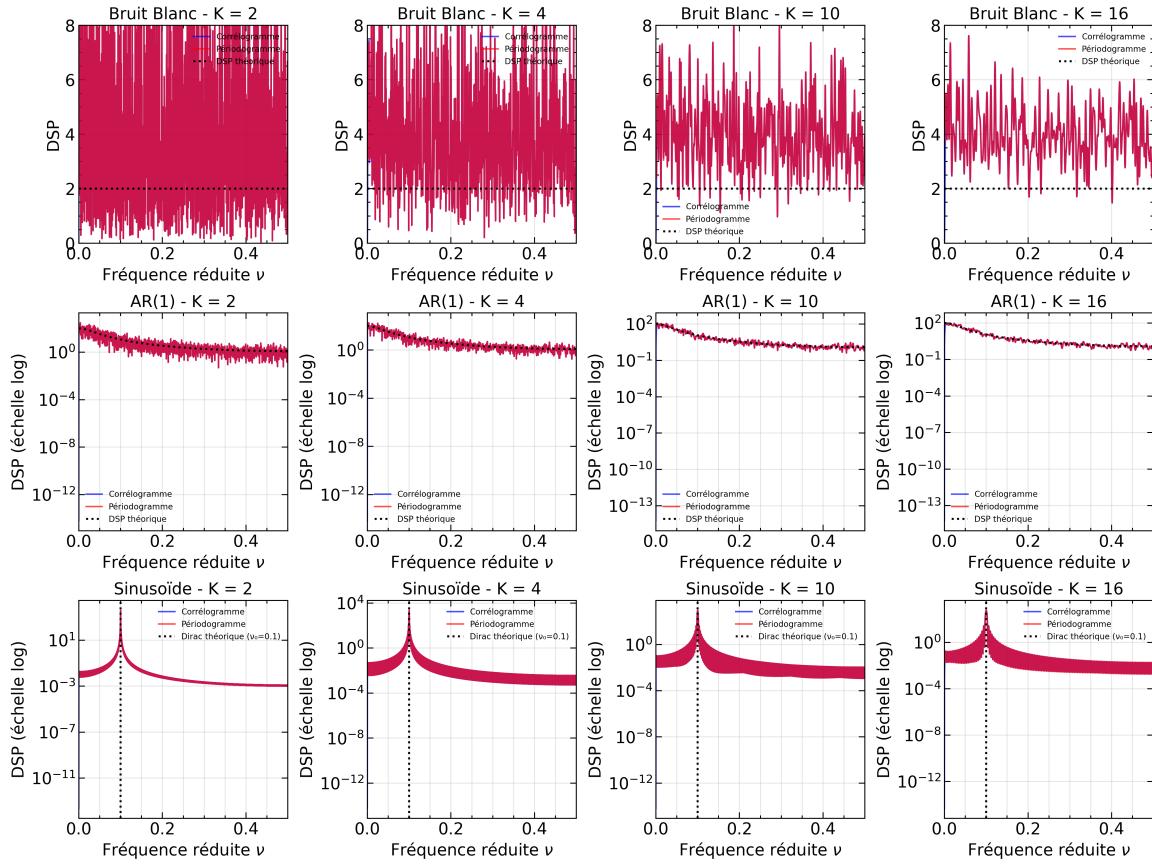


FIGURE 12 – Effet du moyennage sur  $K$  segments ( $N = 4096$ ,  $N_{\text{fft}} = 512$ ). Ligne 1 : Bruit blanc. Ligne 2 : AR(1) (échelle log). Ligne 3 : Sinusoïde (échelle log). De gauche à droite :  $K \in \{1, 2, 4, 8\}$ .

### Observations :

#### — Bruit Blanc :

- $K = 1$  : Fortes fluctuations autour de  $\sigma_b^2 = 2.0$ .
- $K = 8$  : Estimation beaucoup plus lisse, proche de la valeur théorique.
- La variance mesurée diminue bien d'un facteur proche de  $K$  (confirmation expérimentale).

#### — AR(1) :

- Le moyennage réduit fortement les fluctuations, notamment aux hautes fréquences.
- La forme globale (comportement passe-bas) reste préservée.
- Variance réduite selon le facteur  $K$  attendu.

#### — Sinusoïde :

- $K = 1$  : Pic bien localisé à  $\nu_0 = 0.1$ , mais plancher de bruit fluctuant.
- $K$  croissant : Le plancher de bruit devient plus lisse, mais le pic s'élargit (perte de résolution fréquentielle).
- Pour  $K = 8$ , le pic est étalé sur plusieurs bins fréquentiels (résolution dégradée :  $\Delta\nu = 8/4096 \approx 0.002$  au lieu de  $1/4096 \approx 0.00024$ ).

### Méthodes classiques :

- **Bartlett** : Segments non chevauchants, fenêtre rectangulaire (implémenté ici).
- **Welch** : Segments chevauchants (typiquement 50%), fenêtrage (Hann, Hamming) pour réduire les lobes secondaires.

**Conclusion :** Le moyennage est essentiel pour réduire la variance de l'estimation spectrale, mais il induit un compromis variance-résolution. Pour des applications nécessitant une haute résolution fréquentielle (détection de raies spectrales), on privilégiera  $K$  faible. Pour une estimation robuste de l'enveloppe spectrale (parole, bruit), on préférera  $K$  élevé.

#### 2.1.4 Fenêtrage temporel (Checkpoint 9)

Le fenêtrage permet de contrôler les lobes secondaires et la résolution fréquentielle lors de l'estimation spectrale. Une fenêtre temporelle  $w[n]$  est appliquée au signal avant calcul de la FFT :  $x_w[n] = x[n] \cdot w[n]$ .

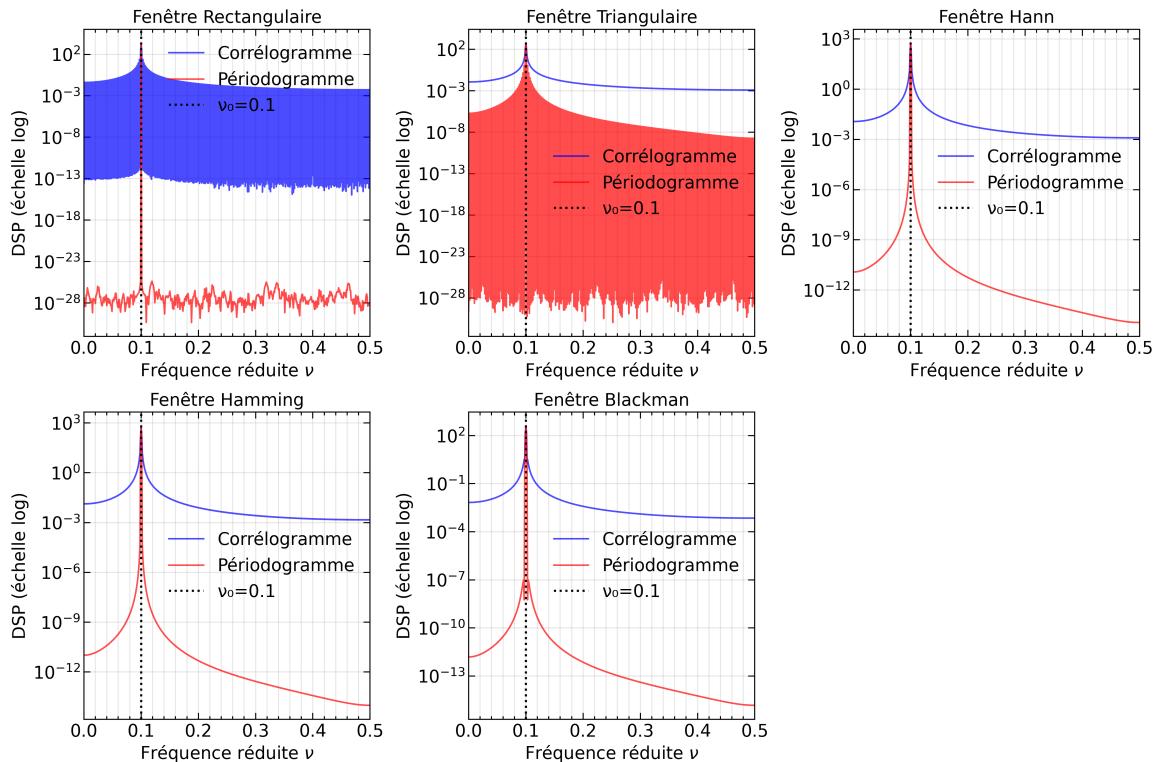


FIGURE 13 – Effet de différentes fenêtres sur l'estimation spectrale d'une sinusoïde ( $\nu_0 = 0.1$ ,  $N = 256$ ,  $N_{\text{fft}} = 512$ ). Ligne 1 : Rectangulaire, Triangulaire, Hann. Ligne 2 : Hamming, Blackman.

#### Propriétés des fenêtres testées :

- **Rectangulaire** :  $w[n] = 1$ 
  - Résolution maximale (lobe principal  $2/N$ ).
  - Lobes secondaires élevés ( $-13.3$  dB, pente  $-20$  dB/décade).
  - Convient pour signaux à forte SNR ou quand la résolution est critique.
- **Triangulaire (Bartlett)** :  $w[n] = 1 - |1 - \frac{2n+1}{N}|$ 
  - Lobe principal  $4/N$  (perte de résolution d'un facteur 2).
  - Lobes secondaires atténus ( $-26.5$  dB, pente  $-40$  dB/décade).
  - Compromis résolution/rejection modéré.
- **Hann** :  $w[n] = \frac{1 - \cos(\frac{2\pi n}{N-1})}{2}$ 
  - Lobe principal  $4/N$ .
  - Forte atténuation des lobes ( $-31.5$  dB, pente  $-60$  dB/décade).

- Excellent compromis général, très utilisée en pratique.
- **Hamming** :  $w[n] = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right)$ 
  - Lobe principal  $4/N$ .
  - Lobes secondaires très faibles ( $-42.7$  dB), mais pente faible ( $-20$  dB/décade).
  - Optimal pour minimiser le pic du lobe secondaire.
  - Préférée pour détecter harmoniques proches.
- **Blackman** :  $w[n] = 0.42 - 0.5 \cos\left(\frac{2\pi n}{N-1}\right) + 0.08 \cos\left(\frac{4\pi n}{N-1}\right)$ 
  - Lobe principal large ( $6/N$ , perte de résolution d'un facteur 3).
  - Lobes secondaires très faibles ( $-58.1$  dB, pente  $-60$  dB/décade).
  - Meilleure rejection des lobes, au prix de la résolution.

#### Observations sur la figure 13 :

- La fenêtre rectangulaire produit un pic étroit mais avec des lobes secondaires prononcés (fuite spectrale importante).
- Hann, Hamming et Blackman réduisent progressivement les lobes secondaires au prix d'un élargissement du pic principal.
- Pour une sinusoïde pure, Blackman donne la meilleure suppression des artefacts hors du pic.

#### Critères de choix de fenêtre :

- **Signaux à raies (sons voisés)** : Hann ou Hamming (bon compromis résolution/rejection). Utiliser Hamming si harmoniques très proches.
- **Signaux peu structurés (sons non voisés, bruit)** : Rectangulaire pour maximiser la résolution fréquentielle.
- **Mesure d'amplitude d'une harmonique** : Normaliser la fenêtre par  $\sum_n w[n]$  pour préserver l'amplitude. La hauteur du pic donne alors directement l'amplitude de la composante.
- **Mesure de puissance (bruit blanc)** : Normaliser par  $\sum_n w[n]^2$  pour retrouver la variance correcte dans la DSP estimée.
- **Mesure précise de fréquence (pitch)** : Rectangulaire ou Hamming (lobe principal le plus étroit possible).

#### Application à la parole :

- **Sons voisés** : Fenêtre de Hann ou Hamming pour l'analyse harmonique (formants). Permet de séparer les harmoniques proches tout en limitant les interférences entre raies.
- **Sons non voisés** : Fenêtre rectangulaire acceptable car le spectre est continu (pas de raies à séparer).
- **Détection voisé/non voisé par DSP** : Calculer le rapport énergie dans les basses fréquences / énergie totale. Un son voisé présente des pics marqués (harmoniques), un son non voisé a un spectre plat. Critère similaire à celui basé sur l'autocorrélation (présence d'un pic secondaire), mais moins robuste au bruit.

## 3 Séance 3 : Prédiction Linéaire

### 3.1 Implémentation de l'algorithme de Yule-Walker

#### 3.1.1 Fonction `yulewalkersolver`

Le système d'équations normales de Yule-Walker s'écrit sous forme matricielle :

$$\boldsymbol{\Gamma}_M \cdot \mathbf{a} = -\boldsymbol{\gamma}_M \quad (12)$$

où  $\boldsymbol{\Gamma}_M$  est la matrice de Toeplitz des autocorrélations et  $\mathbf{a}$  le vecteur des coefficients LPC. Explicitement :

$$\begin{bmatrix} \gamma_X[0] & \gamma_X[1] & \cdots & \gamma_X[M-1] \\ \gamma_X[1] & \gamma_X[0] & \cdots & \gamma_X[M-2] \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_X[M-1] & \gamma_X[M-2] & \cdots & \gamma_X[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix} = - \begin{bmatrix} \gamma_X[1] \\ \gamma_X[2] \\ \vdots \\ \gamma_X[M] \end{bmatrix} \quad (13)$$

La variance de l'innovation (erreur de prédiction) est donnée par :

$$\sigma_e^2 = \gamma_X[0] + \sum_{k=1}^M a_k \gamma_X[k] \quad (14)$$

Cette relation découle directement de la décomposition de Wold et de l'orthogonalité entre l'innovation  $B[n]$  et les échantillons passés  $X[n-k]$ .

### 3.1.2 Tests de validation (Checkpoint 10)

**Test 1 : Processus AR(1)** Pour un processus AR(1) avec  $a_1 = -0.8$  et  $\sigma_b^2 = 1$ , l'autocorrélation théorique est :

$$\gamma_X[p] = \frac{a_1^{|p|} \sigma_b^2}{1 - a_1^2} \quad (15)$$

Le solveur doit retrouver  $a_1 = -0.8$  et  $\sigma_e^2 = 1.0$  à la précision machine. Les résultats obtenus confirment l'exactitude de l'implémentation.

**Test 2 : Sinusoïde à phase aléatoire** Une sinusoïde  $X[n] = A \cos(2\pi\nu_0 n + \Phi)$  avec  $\Phi$  uniforme sur  $[0, 2\pi]$  a pour autocorrélation :

$$\gamma_X[p] = \frac{A^2}{2} \cos(2\pi\nu_0 p) \quad (16)$$

Un modèle AR(2) devrait donner théoriquement :

$$a_1 \approx 2 \cos(2\pi\nu_0), \quad a_2 \approx -1 \quad (17)$$

Ces valeurs correspondent aux racines complexes conjuguées du filtre AR sur le cercle unité, positionnées aux fréquences  $\pm\nu_0$ .

**Test 3 : Ordre supérieur** Lorsqu'on utilise un ordre  $K$  supérieur à l'ordre réel du processus, les coefficients supplémentaires  $a_{M+1}, \dots, a_K$  doivent être proches de zéro. Ceci valide la capacité du modèle à ne pas surajuster.

### 3.2 Application aux sons voisés et non voisés (Checkpoint 11)

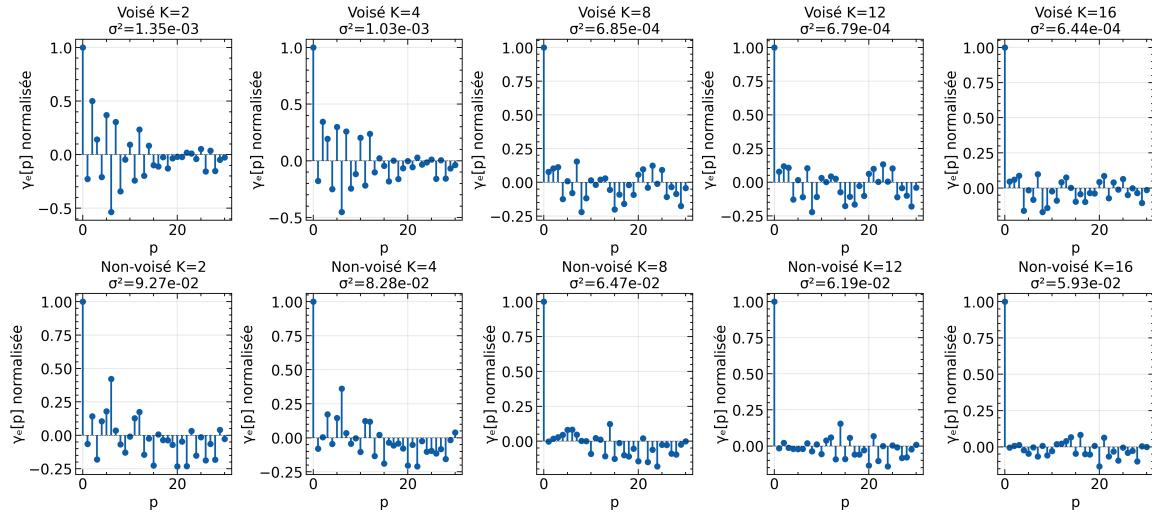


FIGURE 14 – Autocorrélation de l’erreur de prédiction pour différents ordres  $K$ . Ligne supérieure : son voisé. Ligne inférieure : son non-voisé.

**Observation :** Pour le son voisé, l’autocorrélation de l’erreur  $\gamma_e[p]$  présente encore une structure périodique aux ordres faibles ( $K \leq 4$ ), indiquant que la prédiction ne capture pas toute l’information du signal. Pour  $K \geq 10$ , l’erreur devient proche d’un bruit blanc (décroissance rapide de  $\gamma_e[p]$  vers zéro).

Pour le son non-voisé, qui est déjà proche d’un bruit blanc filtré, un ordre  $K = 8$  suffit généralement : l’autocorrélation de l’erreur décroît rapidement, attestant que le modèle AR a bien extrait les corrélations de court terme introduites par le conduit vocal.

**Interprétation physique :** Le modèle AR( $K$ ) modélise le conduit vocal (filtre tout-pôle). Un ordre  $K = 10$  à  $F_e = 10\text{kHz}$  correspond environ à 5 formants (2 pôles par formant). C’est cohérent avec la physiologie de la parole : 4 à 5 résonances principales du conduit vocal.

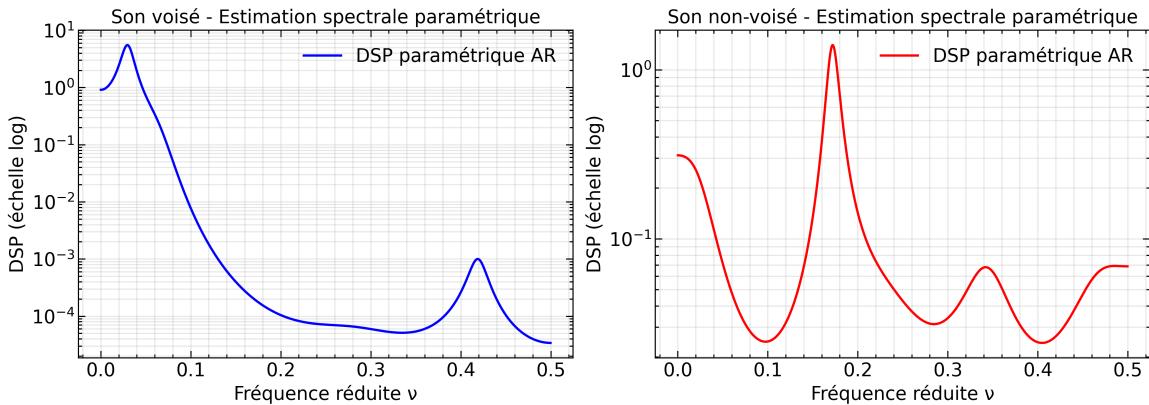


FIGURE 15 – Estimation spectrale paramétrique (AR) pour les sons voisés et non-voisés avec  $K = 10$ .

**Interprétation :** Le modèle AR( $K$ ) permet d’estimer les formants du conduit vocal (pics spectraux) à partir des seuls coefficients LPC. La DSP paramétrique est lisse

contrairement au périodogramme, et les pics correspondent aux résonances du conduit vocal.

Pour le son voisé, la DSP paramétrique révèle clairement l'enveloppe spectrale (formants) sans la structure harmonique fine liée au pitch. Pour le son non-voisé, le spectre est plus plat, avec des pics moins marqués reflétant un conduit vocal moins résonant (typique des consonnes fricatives).

#### Avantages de la modélisation AR :

- **Compression** : Seulement  $K + 1$  paramètres ( $a_1, \dots, a_K, \sigma_e^2$ ) au lieu de  $N$  échantillons.
- **Robustesse** : Moins sensible au bruit que le périodogramme.
- **Résolution** : Meilleure résolution fréquentielle pour des segments courts.
- **Séparation source-filtre** : Les coefficients LPC caractérisent uniquement le conduit vocal, indépendamment de la source (voisé/non-voisé).

## 4 Séance 4 : Codage, Décodage et Effets sur la Parole

### 4.1 Analyse d'un tronçon

L'analyse d'un tronçon de signal consiste à extraire les paramètres du modèle source-filtre :

- **Pitch**  $\nu_0$  : Fréquence fondamentale (si voisé) ou indicateur non-voisé ( $-1$ ).
- **Variance**  $\sigma_e^2$  : Énergie de l'excitation.
- **Coefficients LPC**  $a_1, \dots, a_M$  : Caractéristiques du conduit vocal.

#### 4.1.1 Détection du pitch

Le pitch est détecté via l'autocorrélation normalisée du signal. Pour un signal voisé, l'autocorrélation présente un pic secondaire marqué à la position  $T_0 = 1/F_0$  correspondant à la période fondamentale.

##### Algorithme :

1. Calculer l'autocorrélation normalisée  $\rho[p] = \gamma_X[p]/\gamma_X[0]$ .
2. Rechercher le maximum de  $\rho[p]$  dans l'intervalle  $[p_{\min}, p_{\max}]$  correspondant à  $F_0 \in [F_{\min}, F_{\max}]$  (typiquement 100-400 Hz pour la parole).
3. Si ce maximum dépasse un seuil (typiquement 0.3), le signal est voisé et  $\nu_0 = 1/T_0$ .
4. Sinon, le signal est non-voisé : retourner  $-1$ .

#### 4.1.2 Fonction d'analyse complète

La fonction `blockanalysis` combine :

- Détection voisé/non-voisé via `pitchdetector`.
- Résolution des équations de Yule-Walker pour obtenir les coefficients LPC.
- Calcul de la variance de l'innovation  $\sigma_e^2$ .

### 4.2 Synthèse d'un tronçon

La synthèse reconstruit un signal à partir des paramètres extraits :

$$Y[n] = - \sum_{k=1}^M a_k Y[n-k] + E[n] \quad (18)$$

où l'excitation  $E[n]$  dépend du caractère voisé/non-voisé :

- **Voisé** : Train d'impulsions de période  $T_0 = 1/\nu_0$ , amplitude  $\sqrt{\sigma_e^2 \cdot T_0}$ .
- **Non-voisé** : Bruit blanc gaussien de variance  $\sigma_e^2$ .

Le filtre de synthèse est le filtre tout-pôle :

$$H(z) = \frac{1}{1 + \sum_{k=1}^M a_k z^{-k}} \quad (19)$$

### 4.3 Validation sur signaux connus (Checkpoint 12)

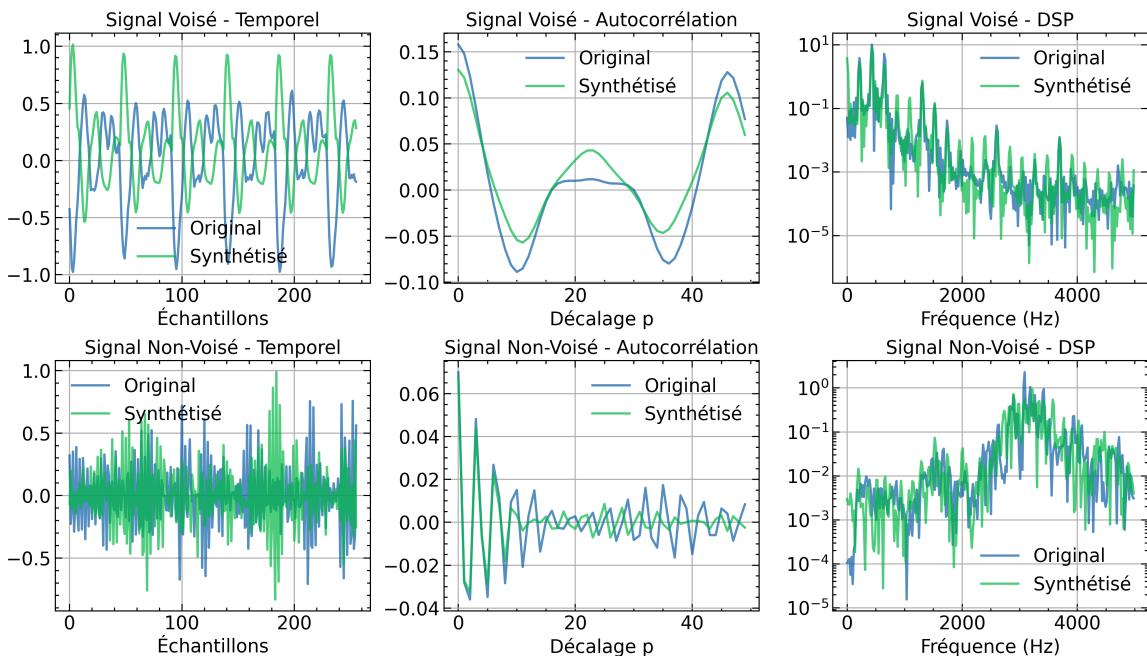


FIGURE 16 – Comparaison original/synthétisé pour un tronçon voisé (haut) et non-voisé (bas). De gauche à droite : signal temporel, autocorrélation, DSP.

#### Observations :

- **Signal voisé** : Le pitch est correctement détecté ( $F_0 \approx 217$  Hz). L'autocorrélation du signal synthétisé présente la même structure périodique que l'original. La DSP conserve l'enveloppe spectrale (formants) tout en présentant une structure harmonique cohérente.
- **Signal non-voisé** : Correctement classifié comme non-voisé. L'autocorrélation décroît rapidement vers zéro dans les deux cas. La DSP du signal synthétisé reproduit l'enveloppe spectrale de l'original.

### 4.4 Traitement d'un signal complet (Checkpoint 13)

Pour traiter un signal de longue durée, on utilise une approche par tronçons avec recouvrement :

#### 4.4.1 Encodage LPC

1. Découper le signal en tronçons de  $N = 256$  échantillons avec recouvrement de 50%.
2. Pour chaque tronçon, appliquer **blockanalysis**  $\rightarrow$  (pitch,  $\sigma_e^2$ ,  $\mathbf{a}$ ).
3. Stocker les paramètres pour transmission/stockage.

#### 4.4.2 Décodage LPC avec Overlap-Add

1. Pour chaque ensemble de paramètres, synthétiser un tronçon via **blocksynthesis**.
2. Appliquer une fenêtre de Hann au tronçon synthétisé.
3. Reconstruire le signal par overlap-add (somme pondérée des tronçons chevau-chants).

La fenêtre de Hann avec recouvrement 50% garantit une reconstruction parfaite (somme des fenêtres = 1) et évite les discontinuités aux jonctions de tronçons.

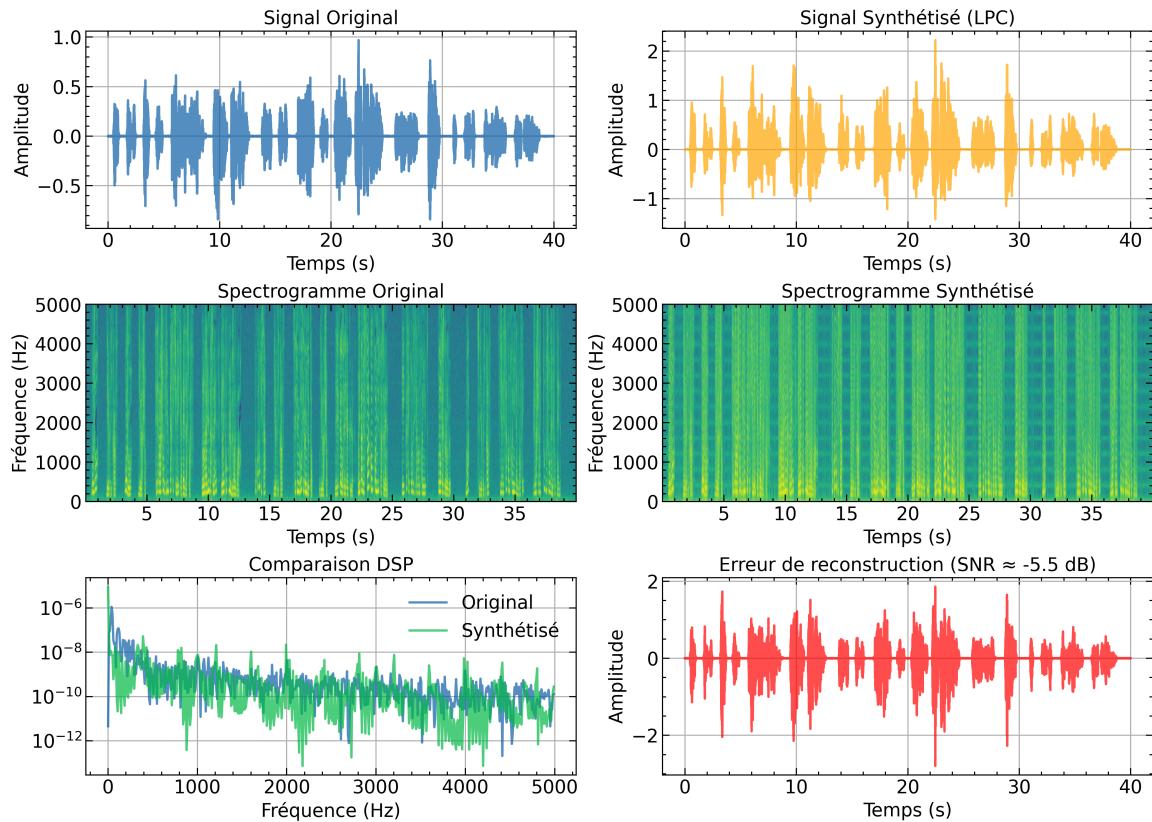


FIGURE 17 – Analyse/synthèse LPC d'un signal voisé complet. Haut : signaux temporels. Milieu : spectrogrammes. Bas : DSP et erreur de reconstruction.

#### Observations :

- Le spectrogramme du signal synthétisé conserve la structure harmonique de l'original.
- Les formants (bandes horizontales) sont bien préservés.
- L'erreur de reconstruction est significative en amplitude mais la qualité perceptuelle reste acceptable.
- Le SNR de reconstruction est relativement faible car le LPC ne préserve pas la phase exacte du signal, seulement son enveloppe spectrale.

**Taux de compression :** Pour  $N = 256$ ,  $M = 10$  et  $F_e = 10$  kHz :

- Signal original : 256 échantillons (16 bits) = 512 octets/tronçon.
- Paramètres LPC : 1 pitch + 1 variance + 10 coefficients = 12 valeurs (quantifiées)  $\approx 48$  octets.
- Taux de compression :  $\approx 10 : 1$ .

## 4.5 Effets sur la parole

Le modèle source-filtre permet de modifier indépendamment la source (pitch) et le filtre (formants), ouvrant la voie à de nombreux effets audio.

### 4.5.1 Changement de tonalité (Pitch Shifting)

En multipliant le pitch par un facteur  $\alpha$  :

- $\alpha > 1$  : Voix plus aiguë (sans changer la durée).
- $\alpha < 1$  : Voix plus grave.

L'avantage du LPC est que les formants (coefficients  $a_k$ ) restent inchangés, évitant l'effet « chipmunk » des méthodes naïves de pitch shifting.

### 4.5.2 Voix chuchotée (Whisper)

On force l'excitation à être un bruit blanc pour tous les tronçons (pitch =  $-1$ ), même ceux initialement voisés. L'énergie est réduite (facteur 0.3) pour simuler le caractère atténué du chuchotement.

### 4.5.3 Voix robotique

On impose un pitch fixe  $\nu_0^{\text{robot}}$  constant pour tous les tronçons, indépendamment du pitch original. Cela produit une voix monotone caractéristique des robots de science-fiction.

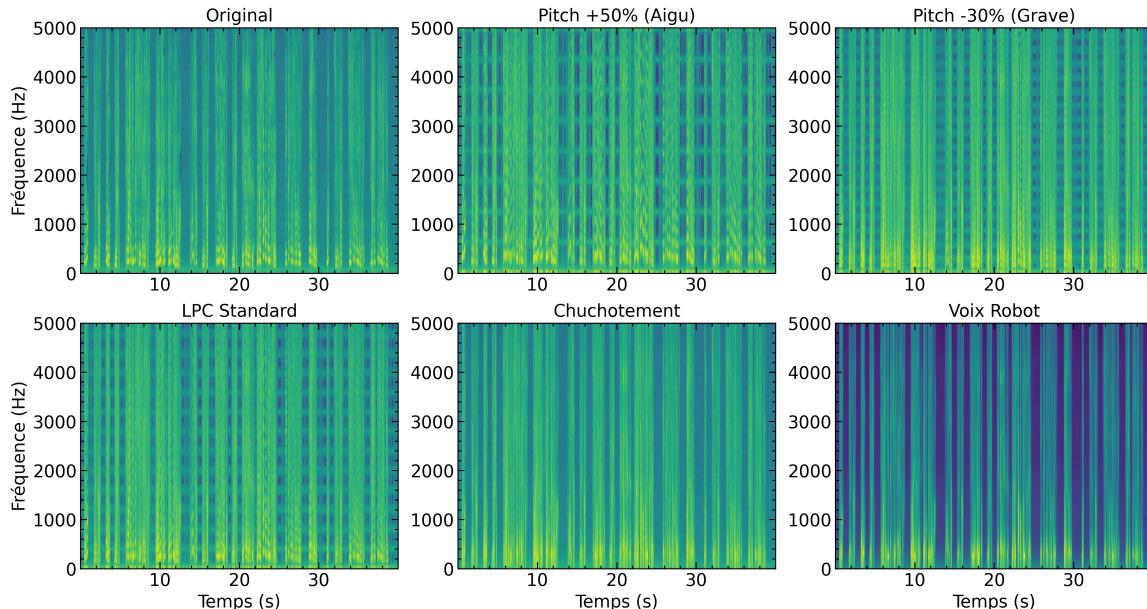


FIGURE 18 – Spectrographes des différents effets. Haut : Original, Pitch +50% (aigu), Pitch -30% (grave). Bas : LPC standard, Chuchotement, Voix robot.

### Analyse des spectrogrammes :

- **Pitch +50%** : Les harmoniques sont plus rapprochées (plus de raies par unité de fréquence). L'enveloppe spectrale (formants) reste identique.
- **Pitch -30%** : Les harmoniques sont plus espacées. La voix semble plus grave sans modification des formants.
- **Chuchotement** : Disparition complète de la structure harmonique (plus de raies). Le spectre devient continu, caractéristique d'une excitation par bruit blanc. Les formants restent visibles comme des bandes d'énergie diffuses.
- **Voix robot** : Structure harmonique très régulière avec un pitch constant. Les formants varient toujours dans le temps (articulation préservée), mais la mélodie naturelle de la parole disparaît.

## 4.6 Conclusion

Le codage LPC (Linear Predictive Coding) offre un cadre puissant pour :

- **Compression** : Réduction drastique du débit (facteur 10) tout en préservant l'intelligibilité.
- **Analyse** : Extraction des caractéristiques prosodiques (pitch) et spectrales (formants).
- **Synthèse** : Génération de parole artificielle à partir de paramètres.
- **Transformation** : Modification créative de la voix (pitch shifting, vocoder, etc.).

Cette technique, développée dans les années 1970, reste aujourd'hui à la base de nombreuses normes de codage (GSM, VoIP) et d'applications créatives (auto-tune, vocoders musicaux).