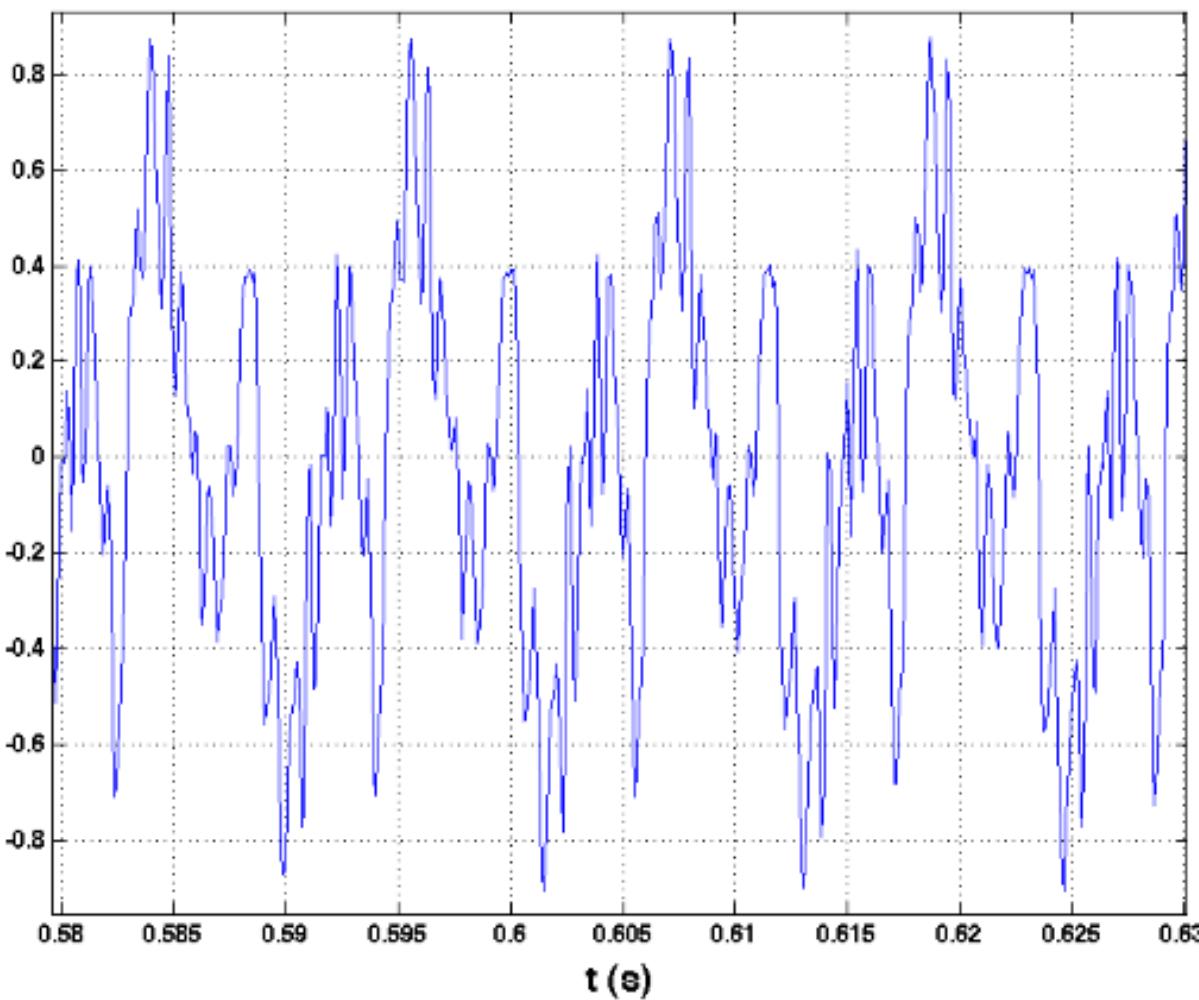


NATHAN LEWY
PAUL LEMAIRE



ANALYSE DE DONNÉES AUDIO

PRÉSENTATION DES RÉSULTATS

TABLE DES MATIÈRES

01 Recherche de fondamentale

- FFT
- Autocorrélation
- Levinson Durbin

02 Séparation de sinus

- Periodogramme
- Levinson Durbin
- Burg
- Musics

03 Analyse d'un signal

- Signal Temporel
 - Fréquences Dominantes
 - Fréquence perçue
- Spectrogramme

04 Conclusion

RECHERCHE DE FONDAMENTALE

RECHERCHE DE FONDAMENTALE

FFT Naive

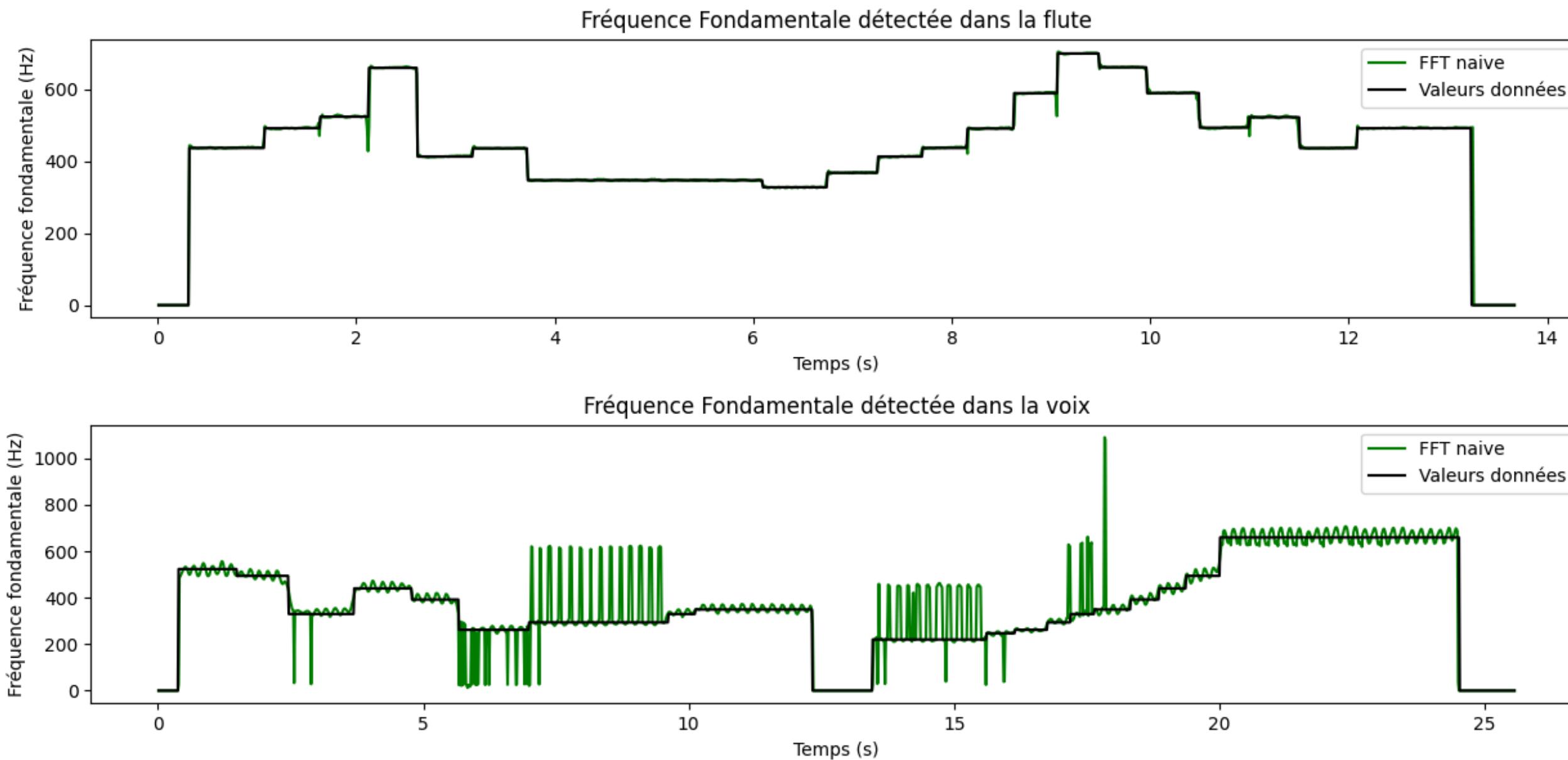
1. Découpage en fenêtres
2. Calcul des ffts avec padding sur chacune **rectangle**
3. Calcul de périodogramme
4. **Premier maximum local avec une certaine amplitude, compris dans la fenêtre de recherche**
5. Bonne classif si erreur normalisée des fréquences < seuil

Algorithme utilisé pour trouver la fondamentale

Il nous est imposé **20 ms** de taille de fenêtre et **10 ms** de taille de saut

RECHERCHE DE FONDAMENTALE

FFT Naive



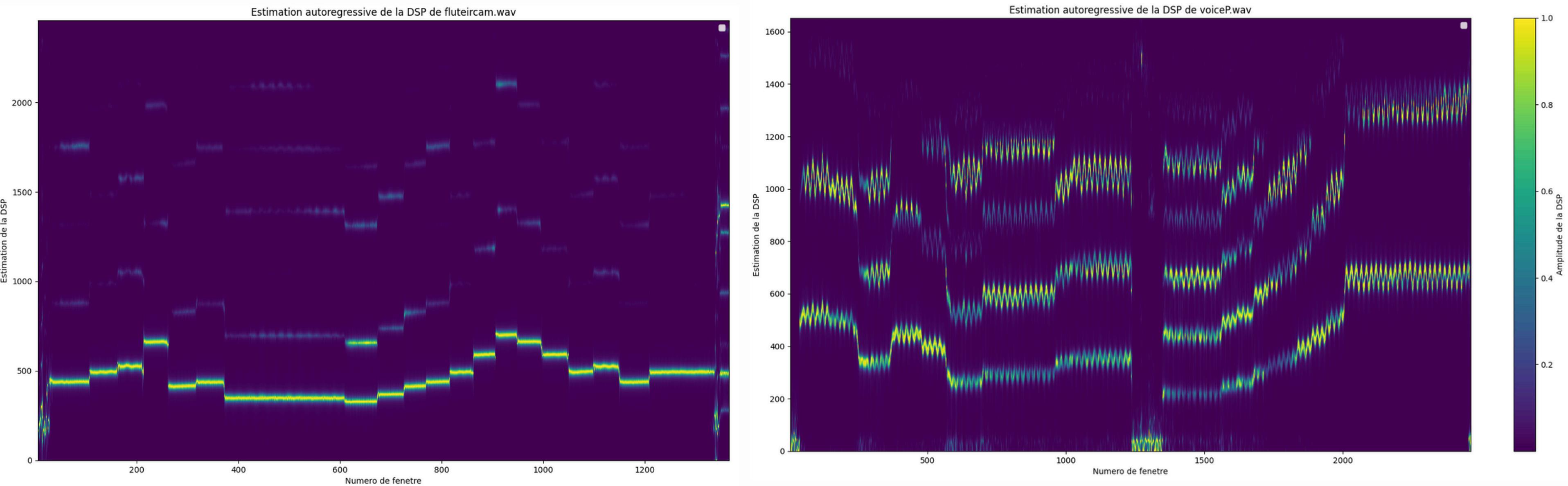
Distance moyenne _p en Hz	$p = 1$	$p = 2$
Flûte	2	34
Voix	74	86

Erreur de 5% on trouve
1279/1367 pour la flute
et **1179/1761** pour la voix

> Fondamentale ou des harmoniques manquantes, ici parce que la fondamentale est plus faible que ses harmoniques

RECHERCHE DE FONDAMENTALE

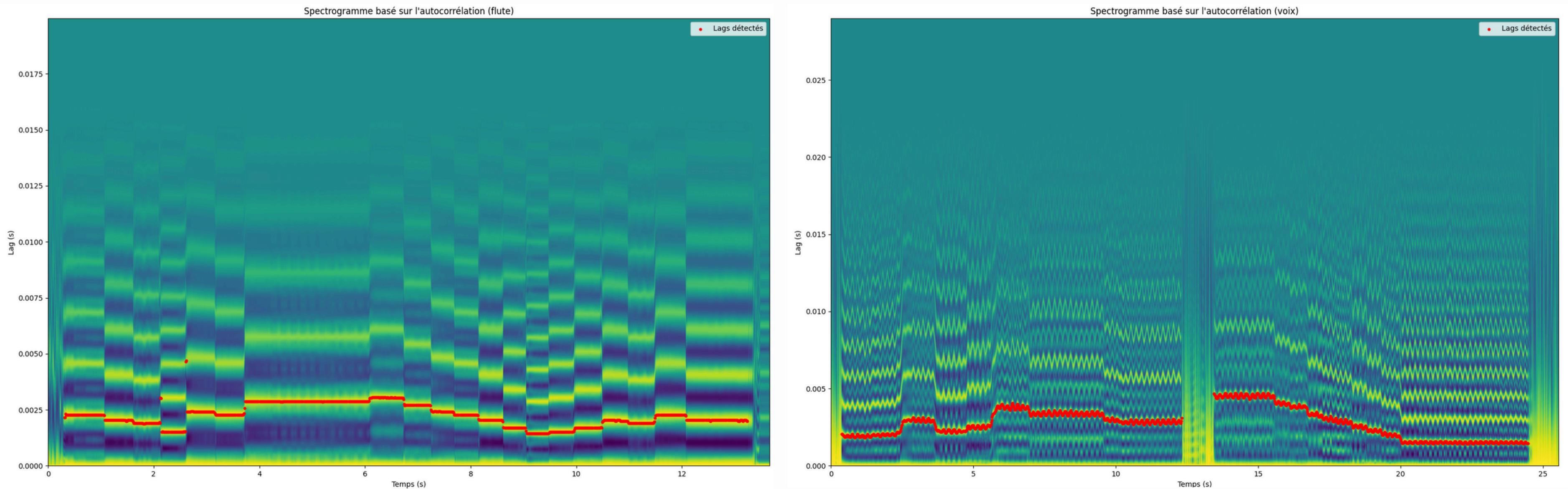
Regard sur les DSP



Estimations en utilisant Levinson Durbin de la DSP des deux signaux

RECHERCHE DE FONDAMENTALE

Regard sur les DSP



Estimations avec le maximum d'autocorrélation

RECHERCHE DE FONDAMENTALE

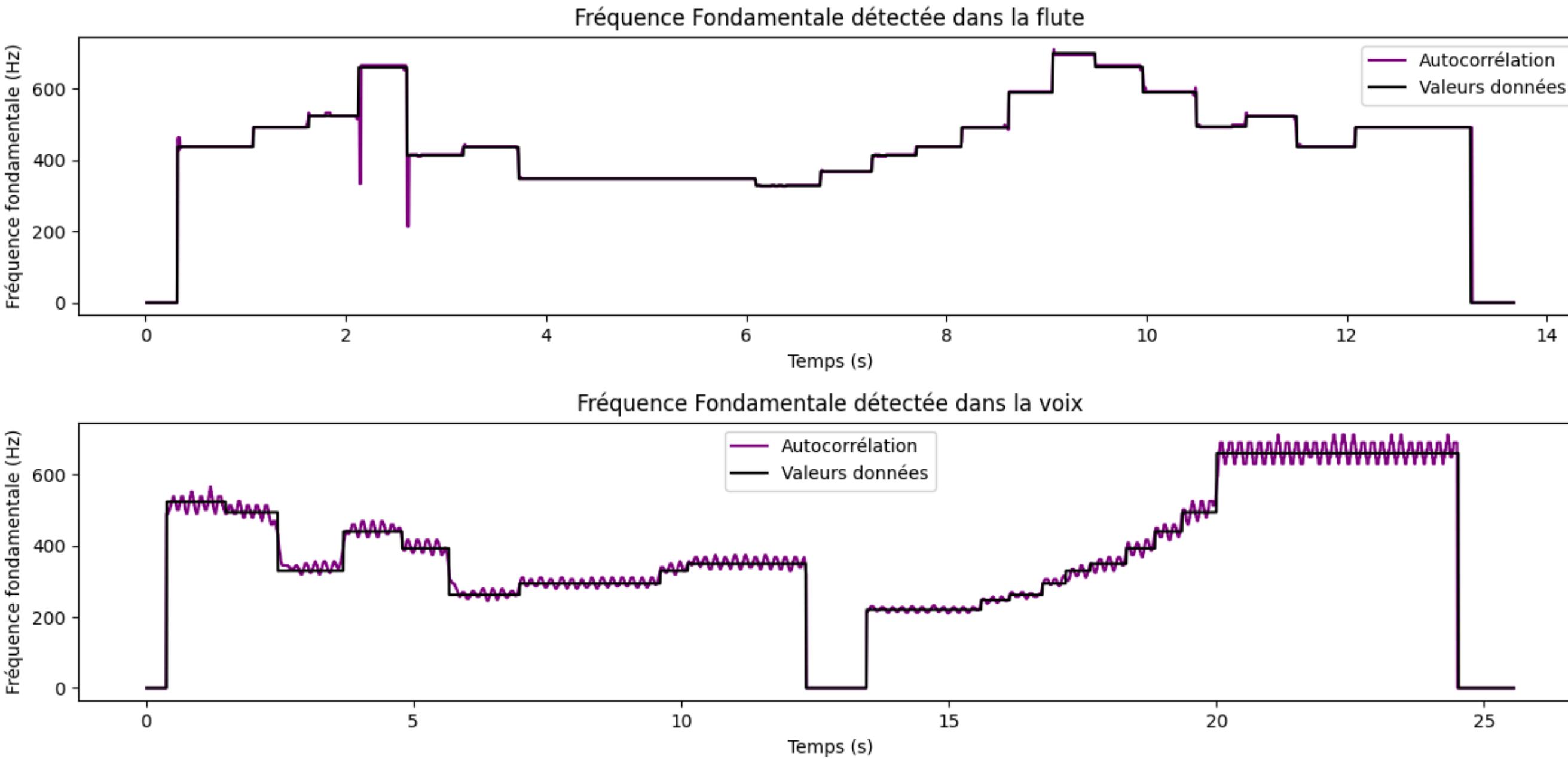
Autocorrélation

1. Découpage en fenêtres
2. Calcul des **R_{xx}** sur chacune, fenêtre **rectangle**
3. Prend l'indice **I_max** du maximum des **R_{xx}**
4. On trouve **f₀ = samplerate / I_max**
5. Bonne classif si erreur normalisée des fréquences < seuil

Algorithme utilisé pour trouver la fondamentale

RECHERCHE DE FONDAMENTALE

Autocorrélation



Distance moyenne p en Hz	$p = 1$	$p = 2$
Flûte	3	21
Voix	12	27

Erreur de 5% on trouve
1281/1367 pour la flute et
1344/1761 pour la voix

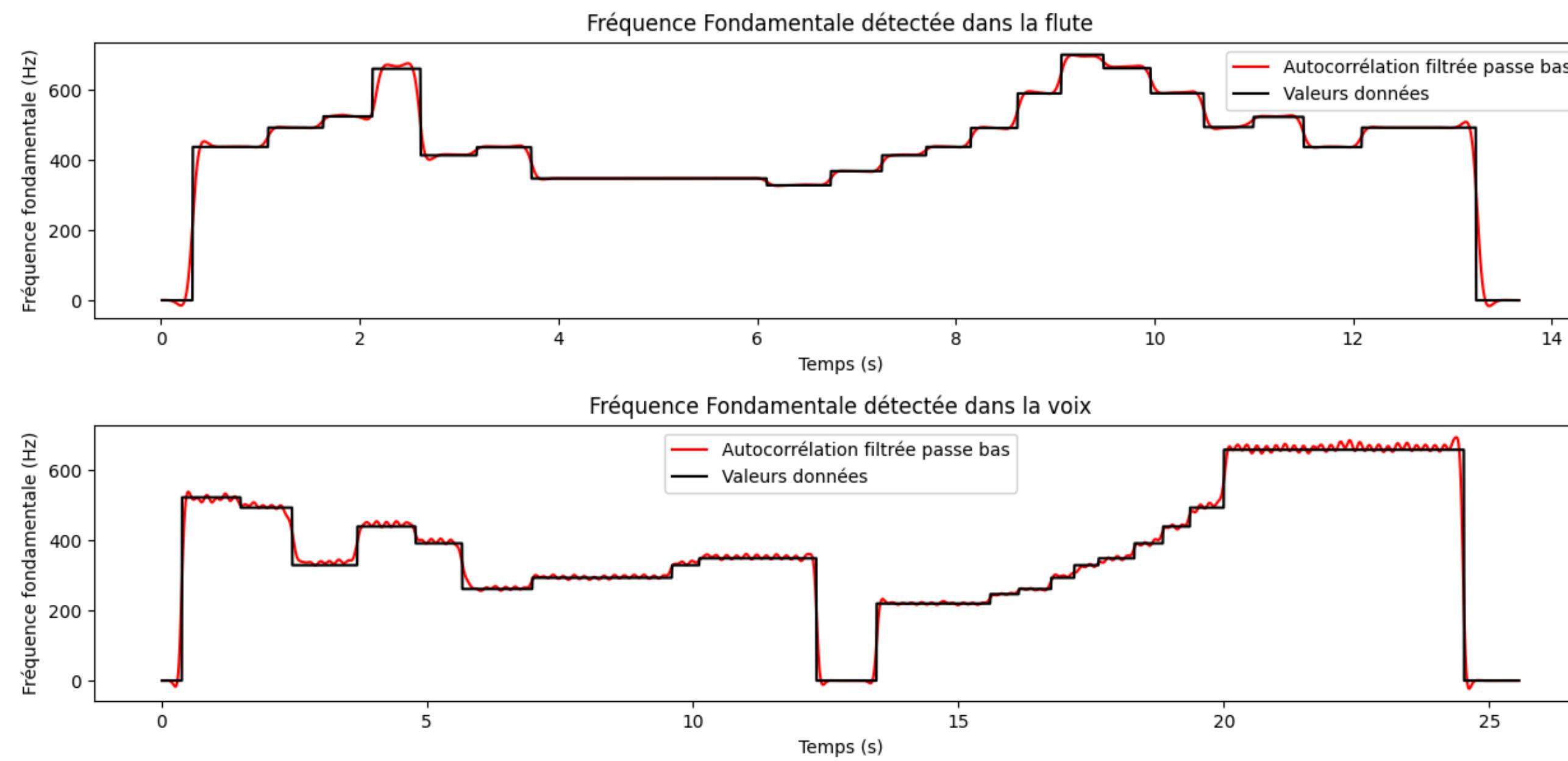
RECHERCHE DE FONDAMENTALE

Autocorrélation filtrée

On ajoute un filtre passe bas **d'ordre 2** de fréquence de coupure **4.5 Hz** sachant que les vibratos humains sont autour de 2/3 Hz

RECHERCHE DE FONDAMENTALE

Autocorrélation filtrée



Distance moyenne _p en Hz	$p = 1$	$p = 2$
Flûte	6	21
Voix	8	22

Erreur de 5% on trouve
1331/1367 pour la flute
et **1493/1761** pour la voix

SEPARATION DE SINUS

SEPARATION DES SINUS

Méthodes non-paramétrique

Méthode de la STFT :

$$S(t, f) = \left| \sum_{n=0}^{N-1} x(n)w(n-t)e^{-j2\pi fn} \right|$$

N = 2048

Hop Size = 10ms

1. On découpe en trames : $x_k(n) = x(n + k \cdot \text{hop_size})$, $0 \leq n < \text{frame_size}$

2. On calcule la fenêtre de Hamming pour les trames :

$$w(n) = 0.5 - 0.5 \cos \left(\frac{2\pi n}{N-1} \right), \quad 0 \leq n < N$$



3. On calcule la fft pour cette trame :

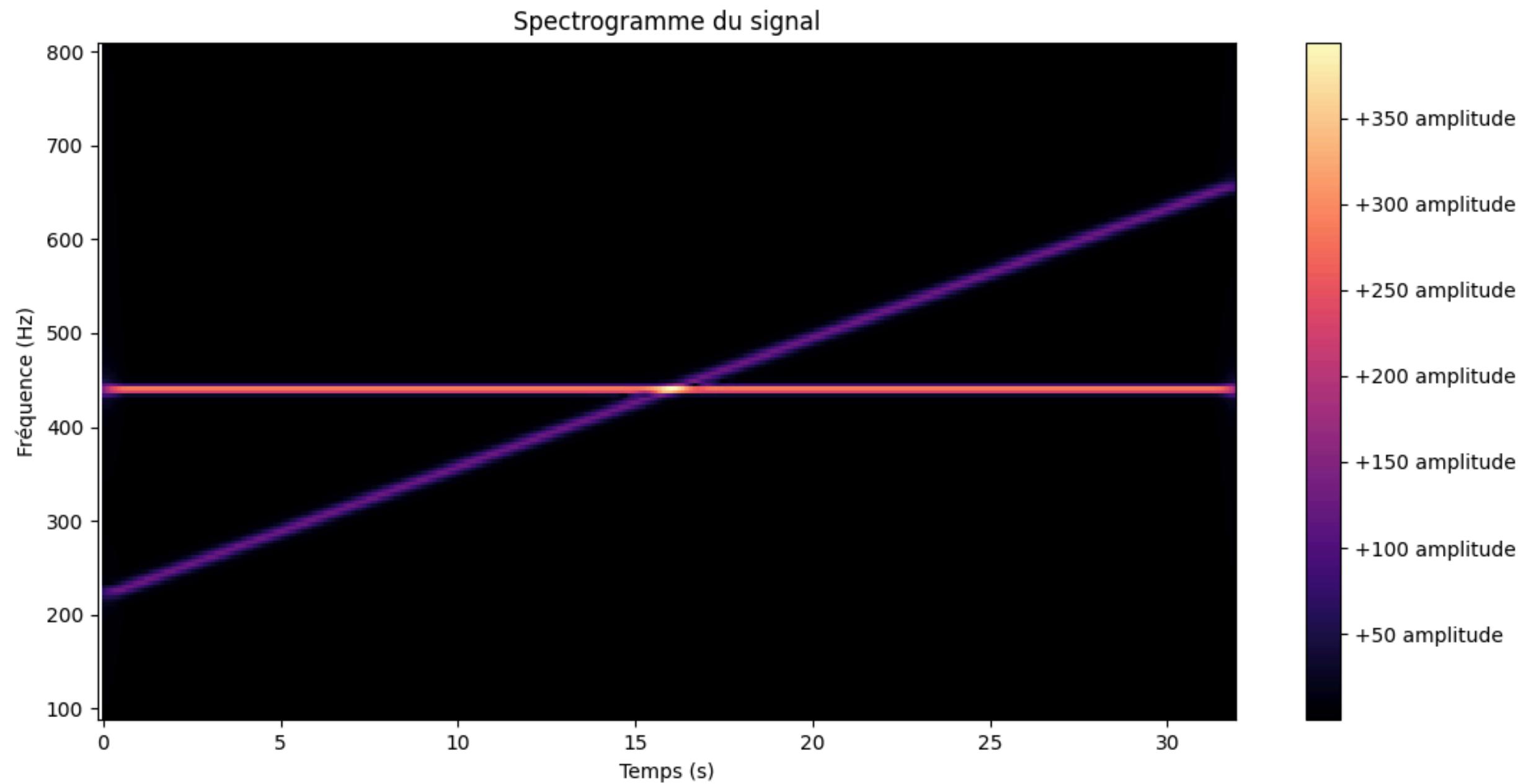
$$X_k(f) = \sum_{n=0}^{N-1} x_k(n)w(n)e^{-j2\pi fn}$$

SEPARATION DES SINUS

Méthodes non-paramétrique

Spectrogramme croisement

- deux sinusoïdes identifiées
- léger bruit au bord



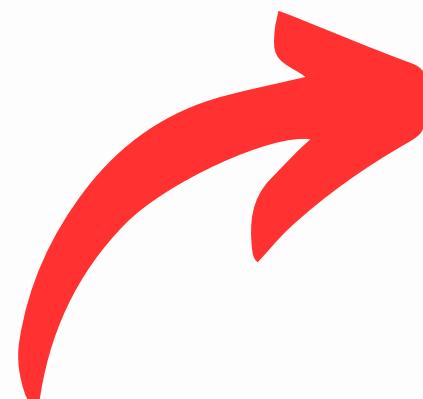
SEPARATION DES SINUS

Méthodes non-paramétrique

Analyse du spectrogramme :

1. Lissage gaussien des amplitudes par fénêtre :

$$S'_m(k) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k'} S_m(k') e^{-\frac{(k-k')^2}{2\sigma^2}}$$



5. Seuillage anti-bruit

$$S'_m(k) > \alpha \max(S'_m)$$

2. Détection des maximaux locaux :

$$S'_m(k - 1) < S'_m(k) > S'_m(k + 1)$$

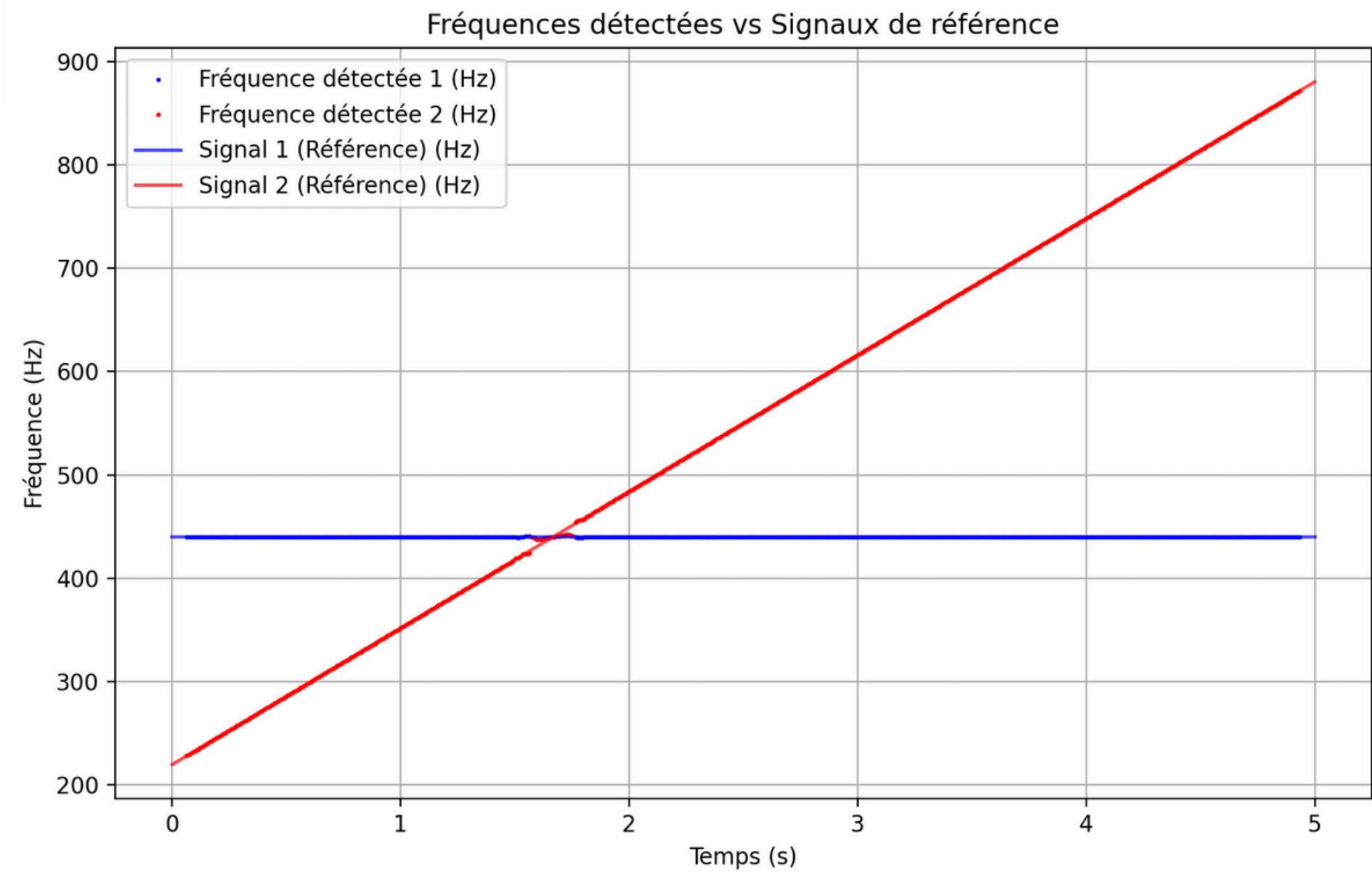
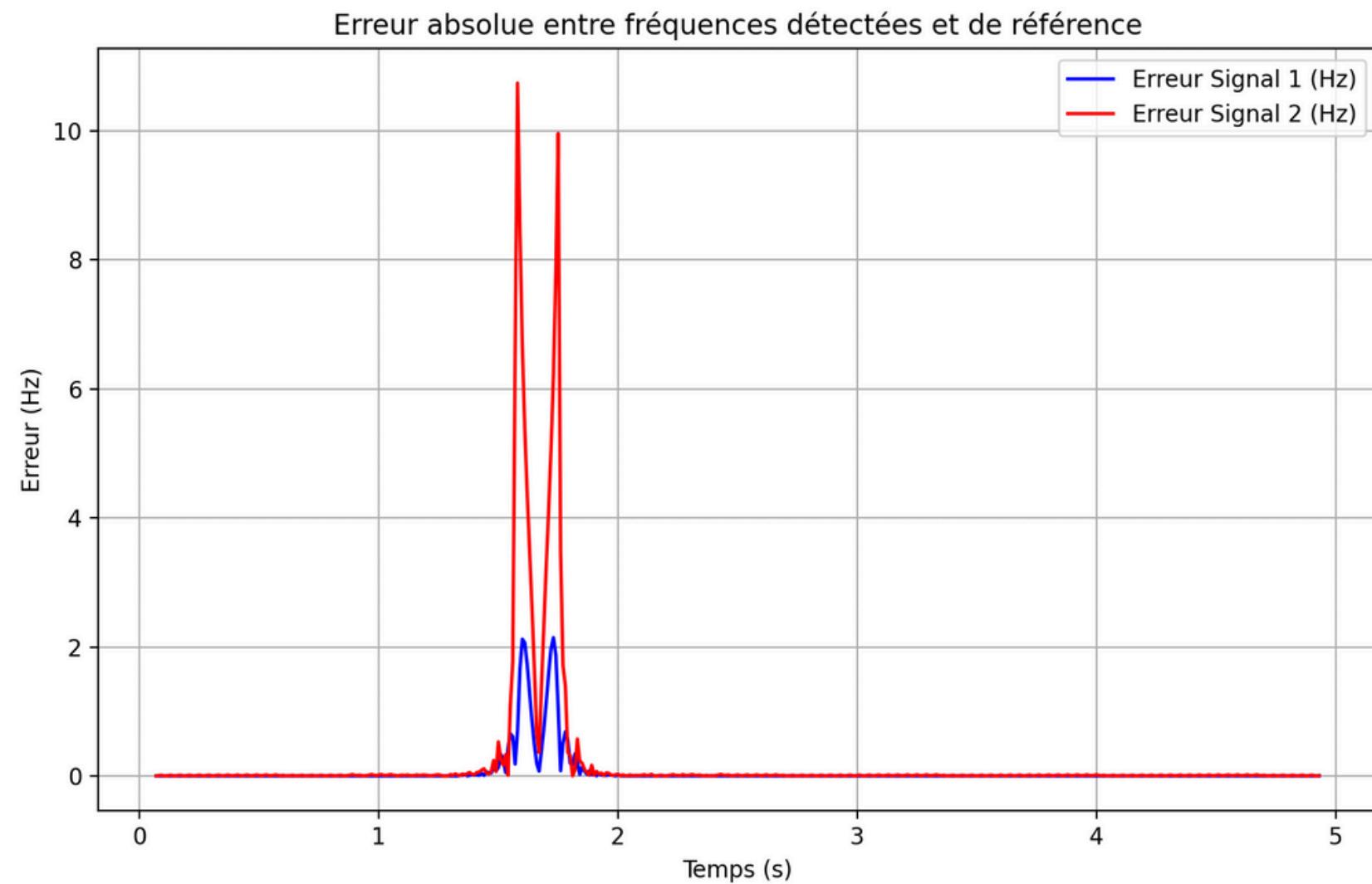
4. Suivi des fréquences

$$\min |f_1(m) - f_1(m - 1)|$$

SEPARATION DES SINUS

Méthodes non-paramétrique

Résultats :



SEPARATION DES SINUS

Méthodes non-paramétrique

Résultats :

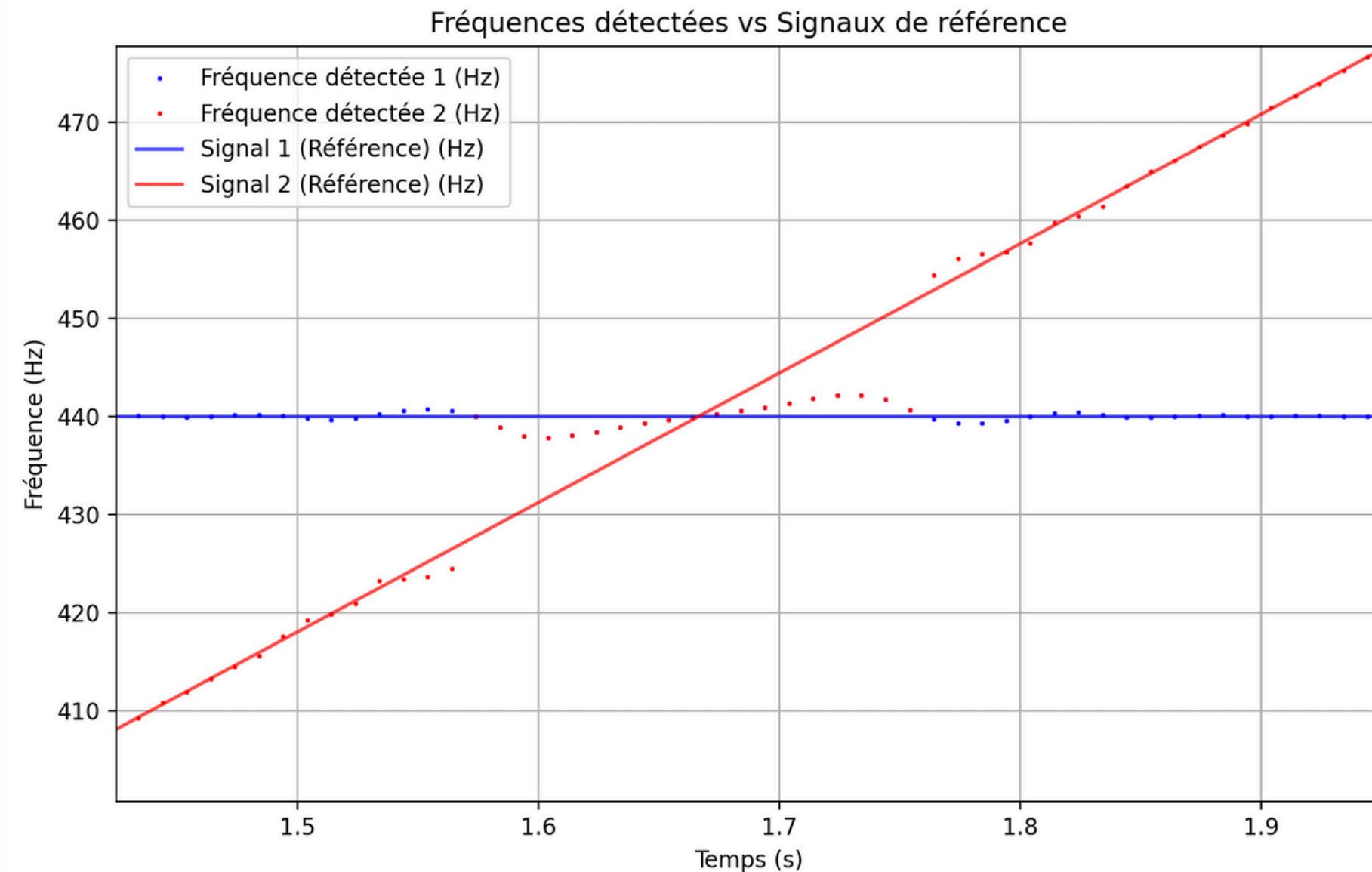
Erreur moyenne pour **Signal 1 : 0.06 Hz**

Erreur moyenne pour **Signal 2 : 0.22 Hz**

Trames avec erreur > 1% pour **Signal 1 : 0/501 (0.00%)**

Trames avec erreur > 1% pour **Signal 2 : 10/501 (2.00%)**

Zoom sur les points problématiques :



SEPARATION DES SINUS

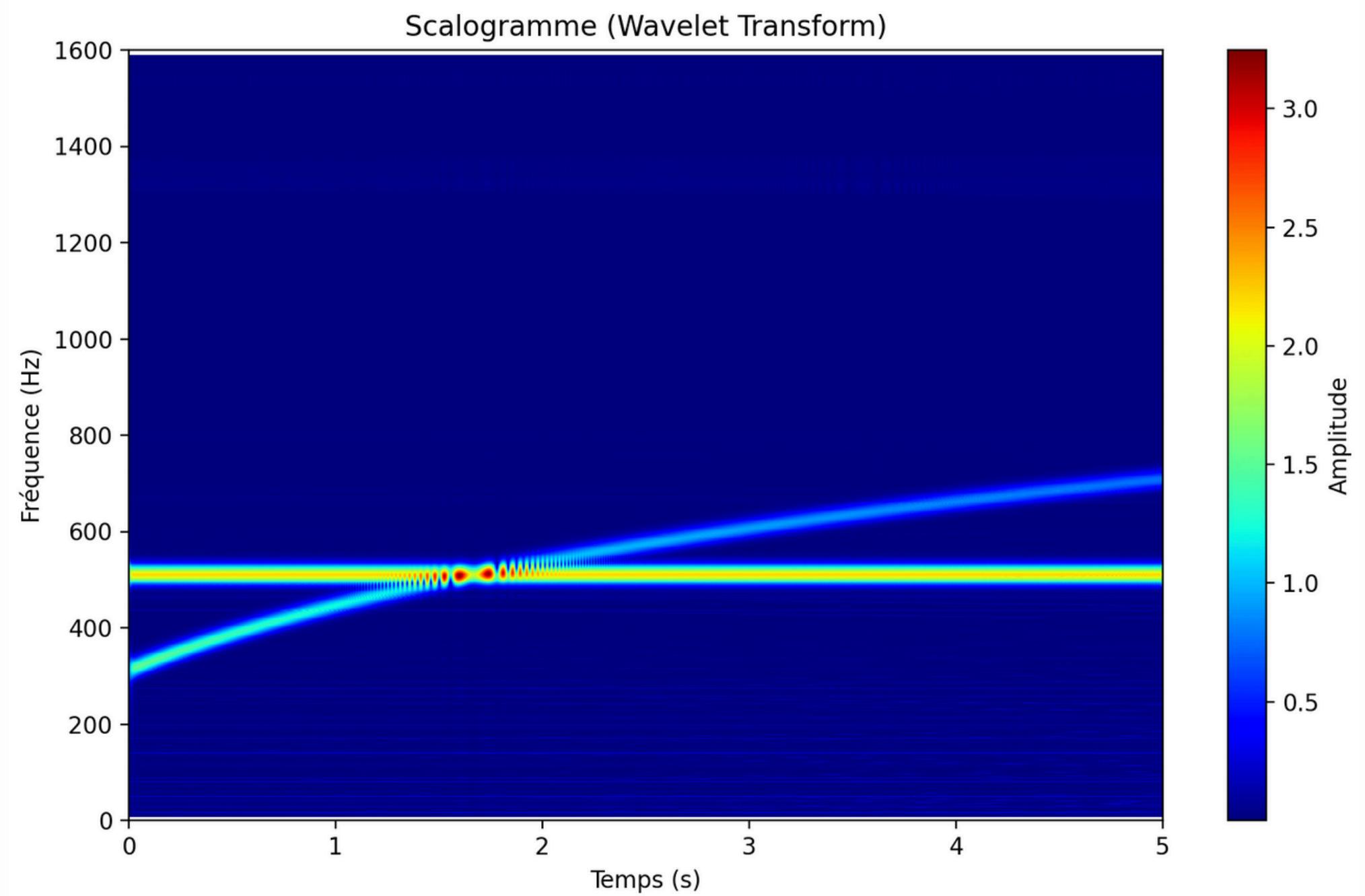
Méthodes non-paramétrique

Autre méthode testée :

Wavelet Transform

Résultats encourageants
mais les problèmes de
résolution fréquentielles sont
plus élevés

Paramètres :
ondelette : Complex Morlet
 $B=2.5$ (temps)
 $C=3.0$ (fréquence)



SEPARATION DES SINUS

Récursion de levinson durbin

$$x_n = \sum_{l=0}^q b_l n_{n-l} - \sum_{k=1}^p a_k x_{n-k}$$

Estimation autorégressive

Formulation matricielle

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(-1) & \dots & R_{xx}(-p) \\ R_{xx}(1) & R_{xx}(0) & \dots & R_{xx}(-(p-1)) \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}(p) & R_{xx}(p-1) & \dots & R_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Resolution en $O(p^3)$

$$P_{xx}(f) = |H(f)|^2 P_n(f) = \frac{\sigma^2}{f_e \left| 1 + \sum_{k=1}^p a_k \exp(-j2\pi kf/f_e) \right|^2}$$

Estimation de la DSP

SEPARATION DES SINUS

Récursion de levinson durbin

1. Découpage en fenêtres
2. Calcul des coeffs d'autocorrélation
3. Resolution du problème matriciel
4. Estimation de la DSP
5. Suivi des fréquences:
 - a. En amplitude si $D(F) >$ seuil
 - b. En fréquence sinon
6. Bonne classif si rapport des amplitudes $>$ seuil

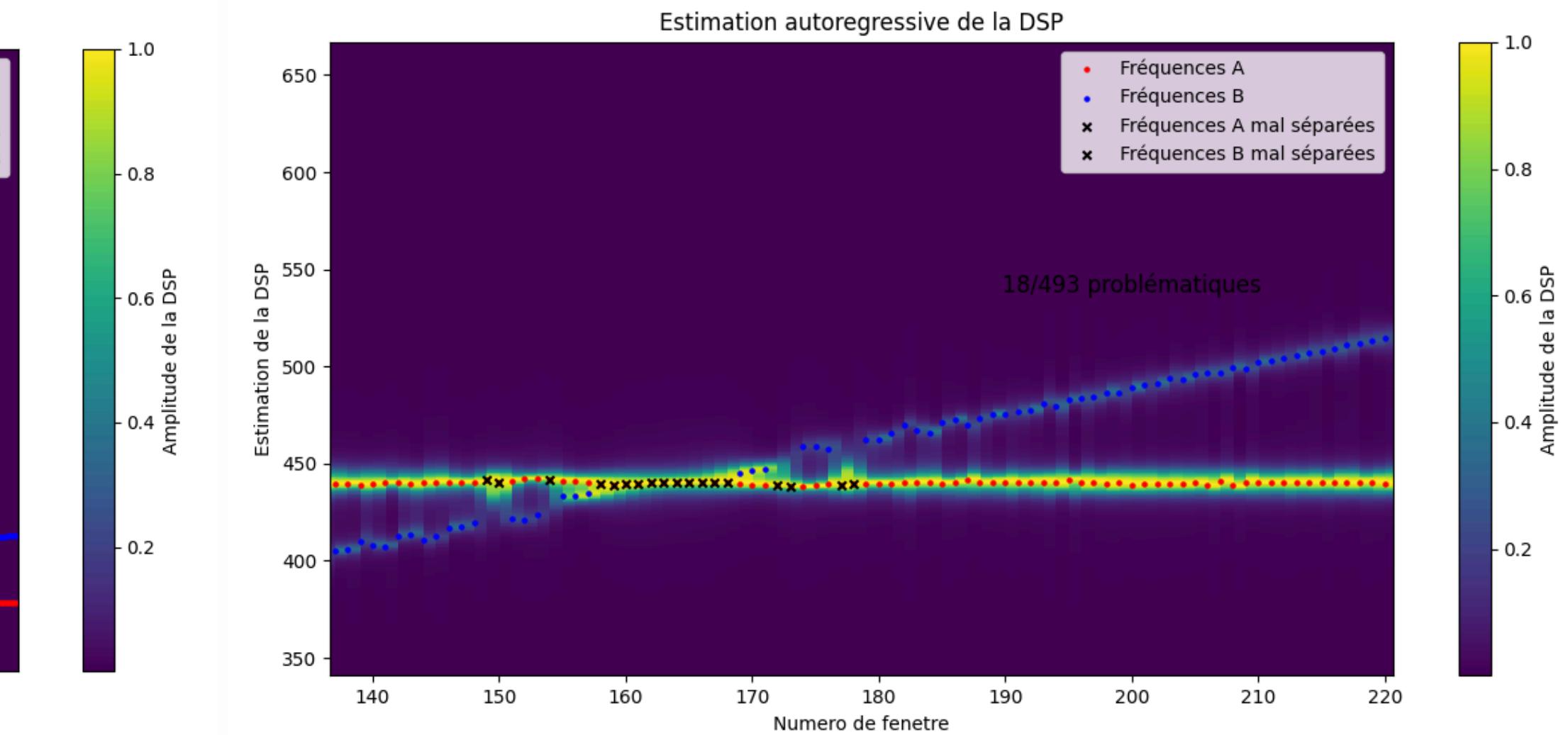
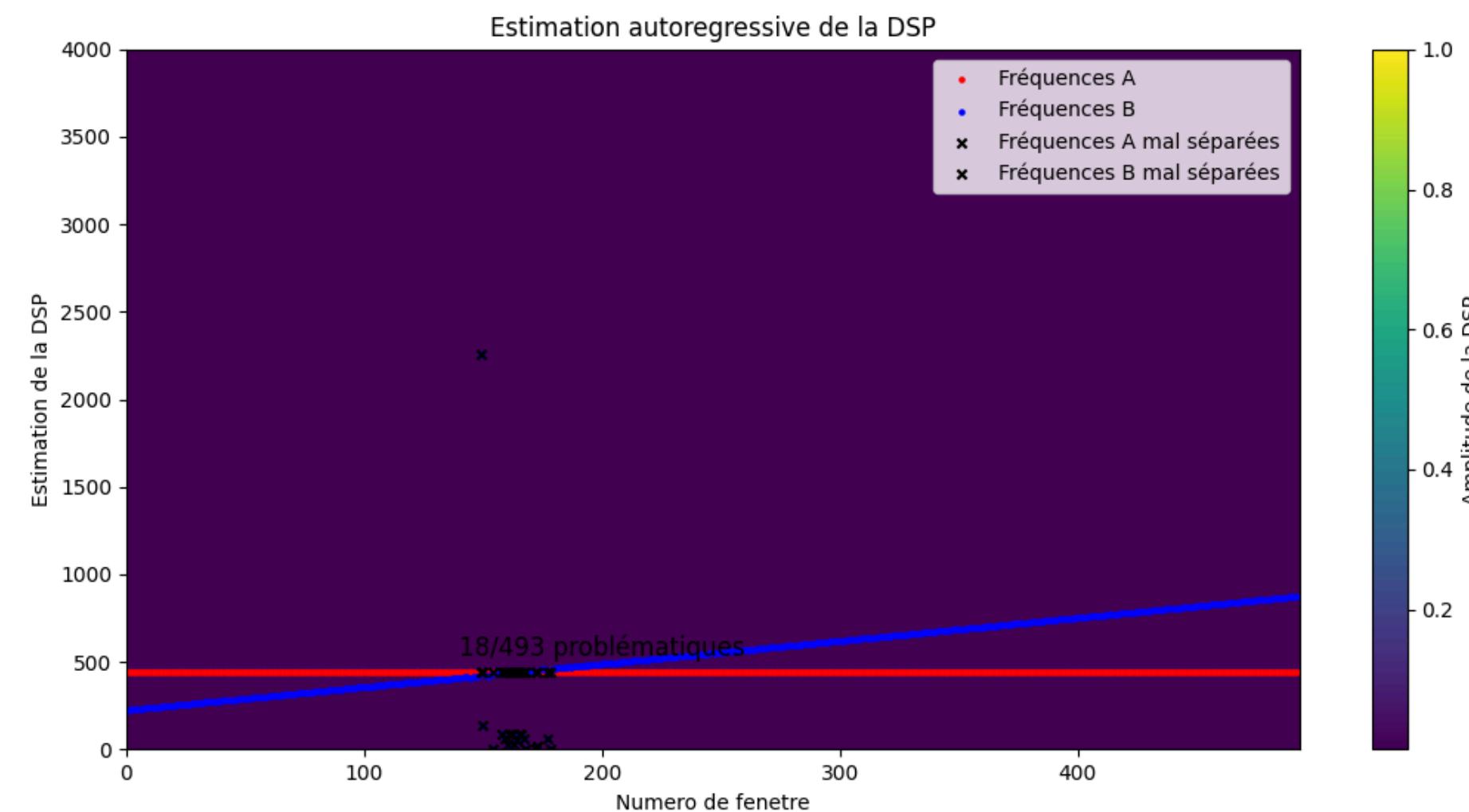
Algorithme utilisé pour séparer

Taille fenêtre = 80 ms
Taille pas = 10 ms
ordre = 300
 $N_{freqs_DSP} = 8000$

Paramètres

SEPARATION DES SINUS

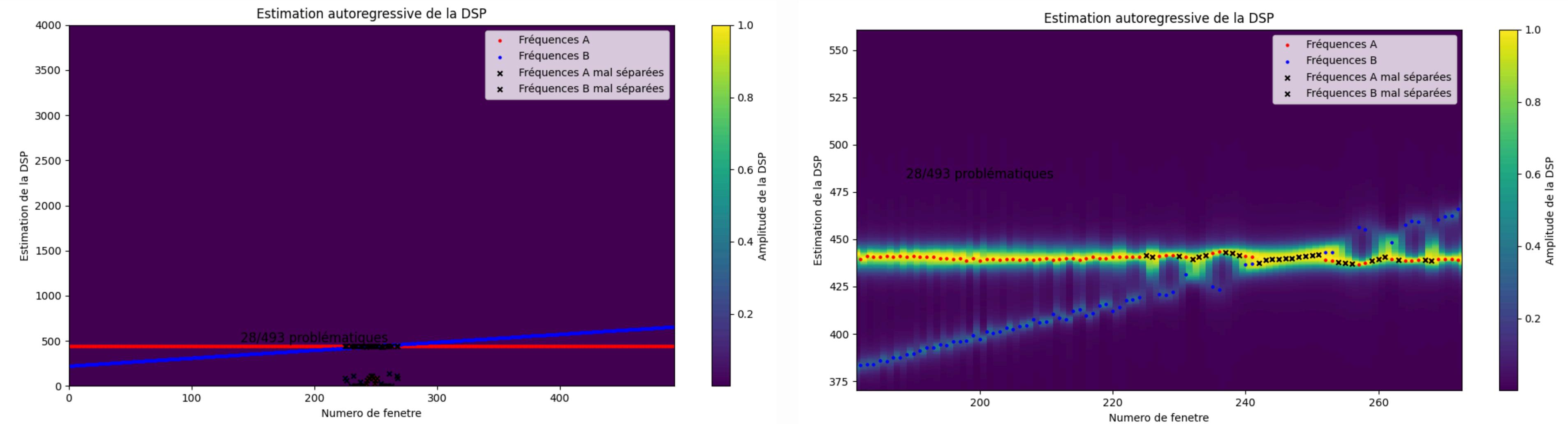
Récursion de levinson durbin



Croisement 1

SEPARATION DES SINUS

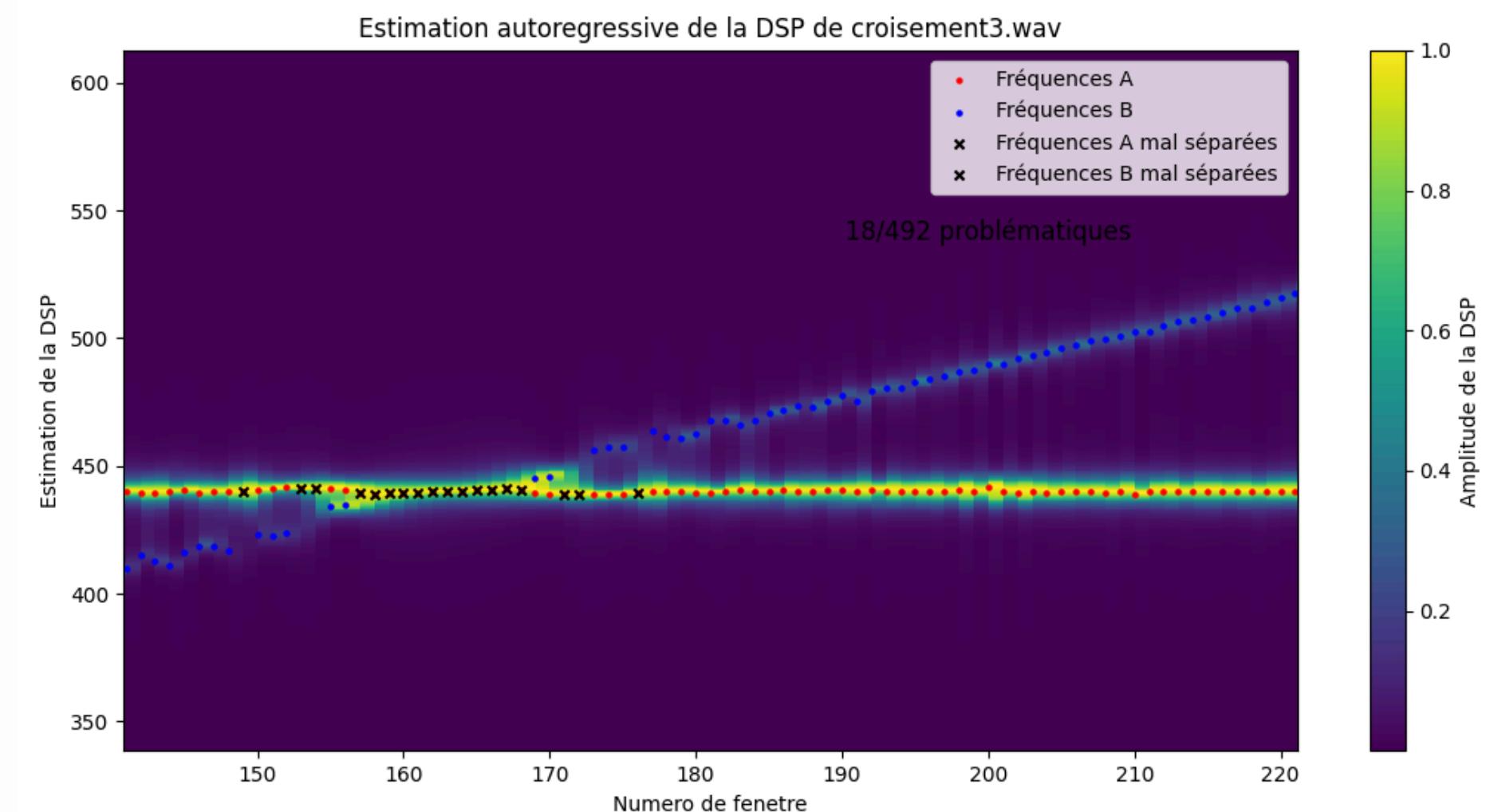
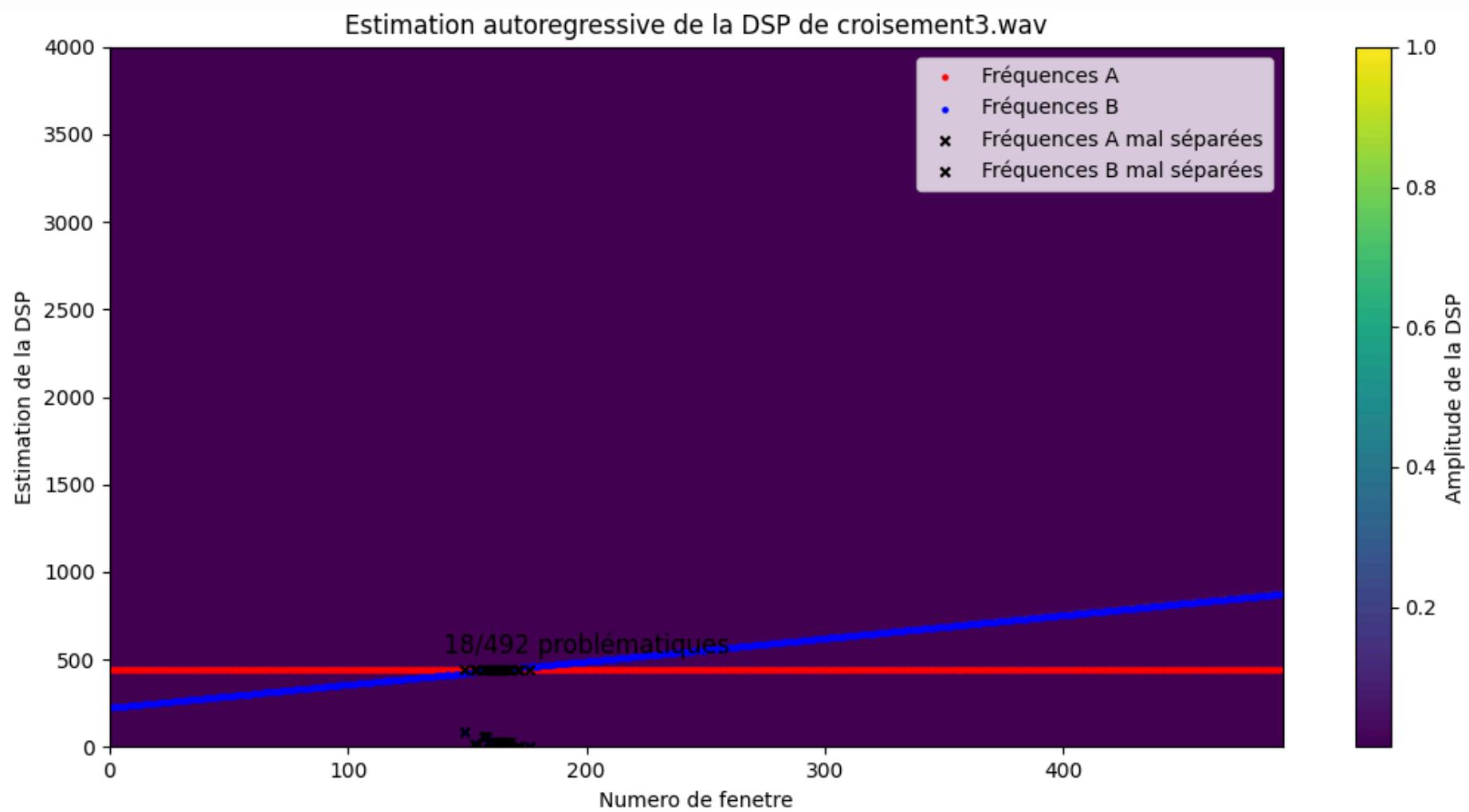
Récursion de levinson durbin



Croisement 2

SEPARATION DES SINUS

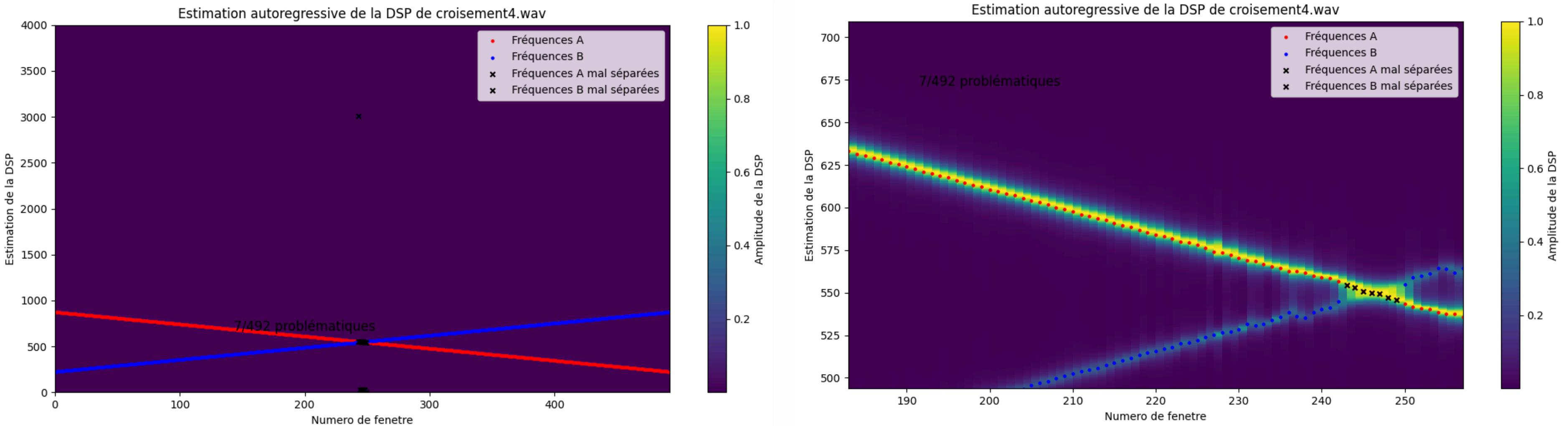
Récursion de levinson durbin



Croisement 3

SEPARATION DES SINUS

Récursion de levinson durbin



Croisement 4

SEPARATION DES SINUS

Méthodes auto-régressives

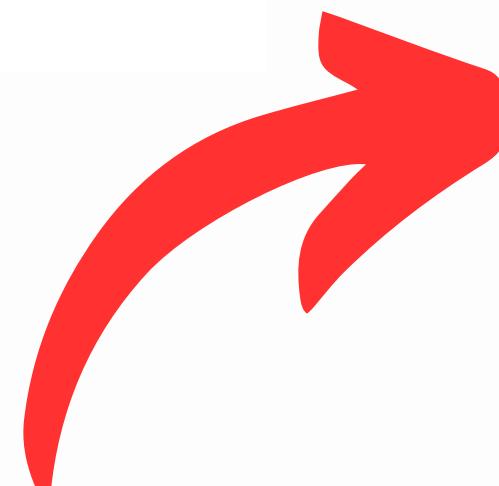
Méthode de burg :

1. Modèle AR :

$$x[n] = - \sum_{k=1}^p a_k x[n-k] + e[n]$$

2. Calcul des coefficients avec la méthode de Burg

$$\sum_{n=p}^{N-1} \left(x[n] + \sum_{k=1}^p a_k x[n-k] \right)^2$$



3. Estimation du spectre de puissance à partir du modèle AR

$$P(f) = \frac{\sigma^2}{|1 + \sum_{k=1}^p a_k e^{-j2\pi kf}|^2}$$

Coefs qui minimisent l'EQ ci dessus

SEPARATION DES SINUS

Méthodes auto-régressives

Spectrogramme : croisement

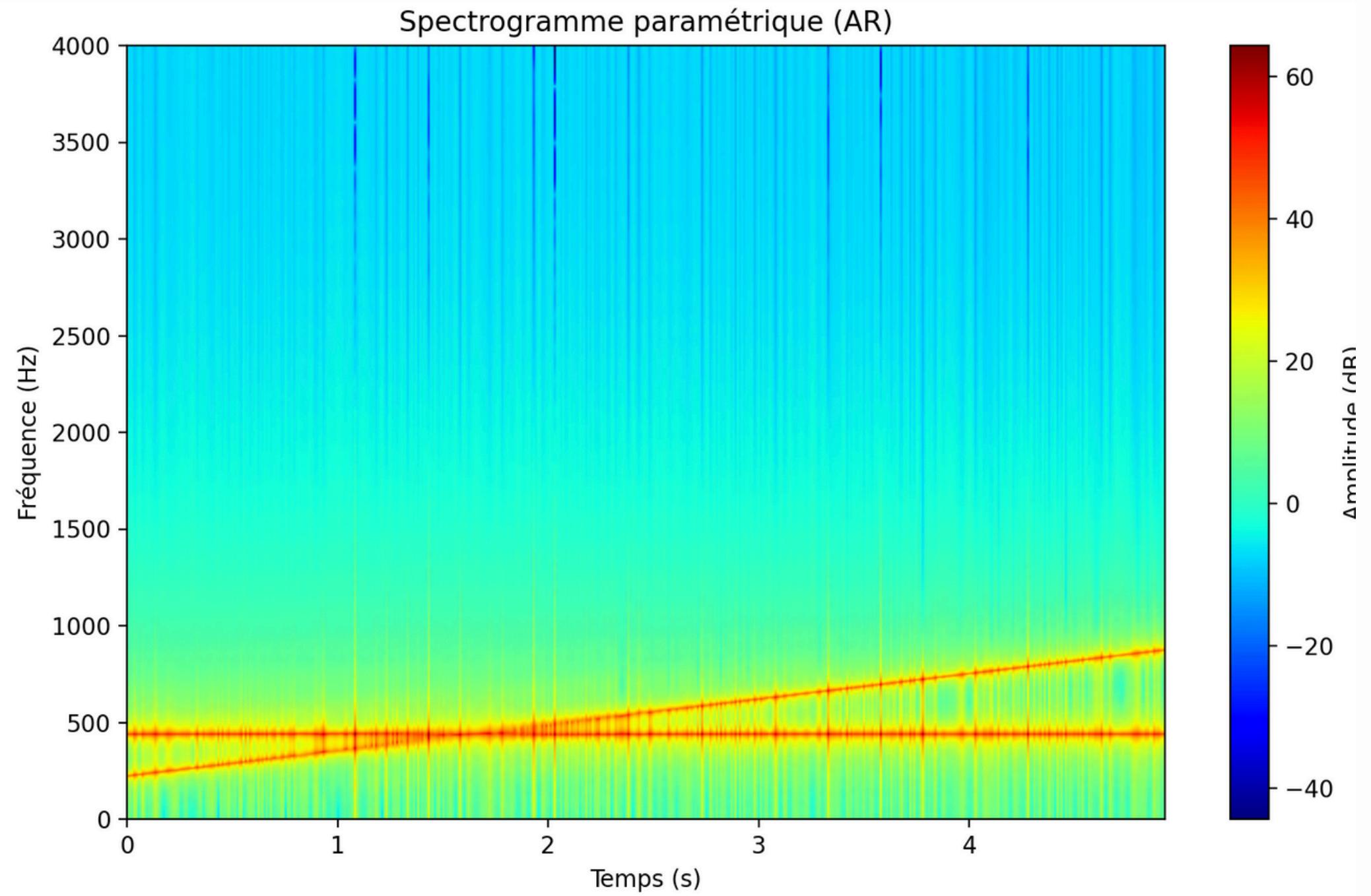
paramètres utilisés :

frame_size = $2^{**}9$

hop_size = 10ms

order = 50

n_fft = $2^{**}15$



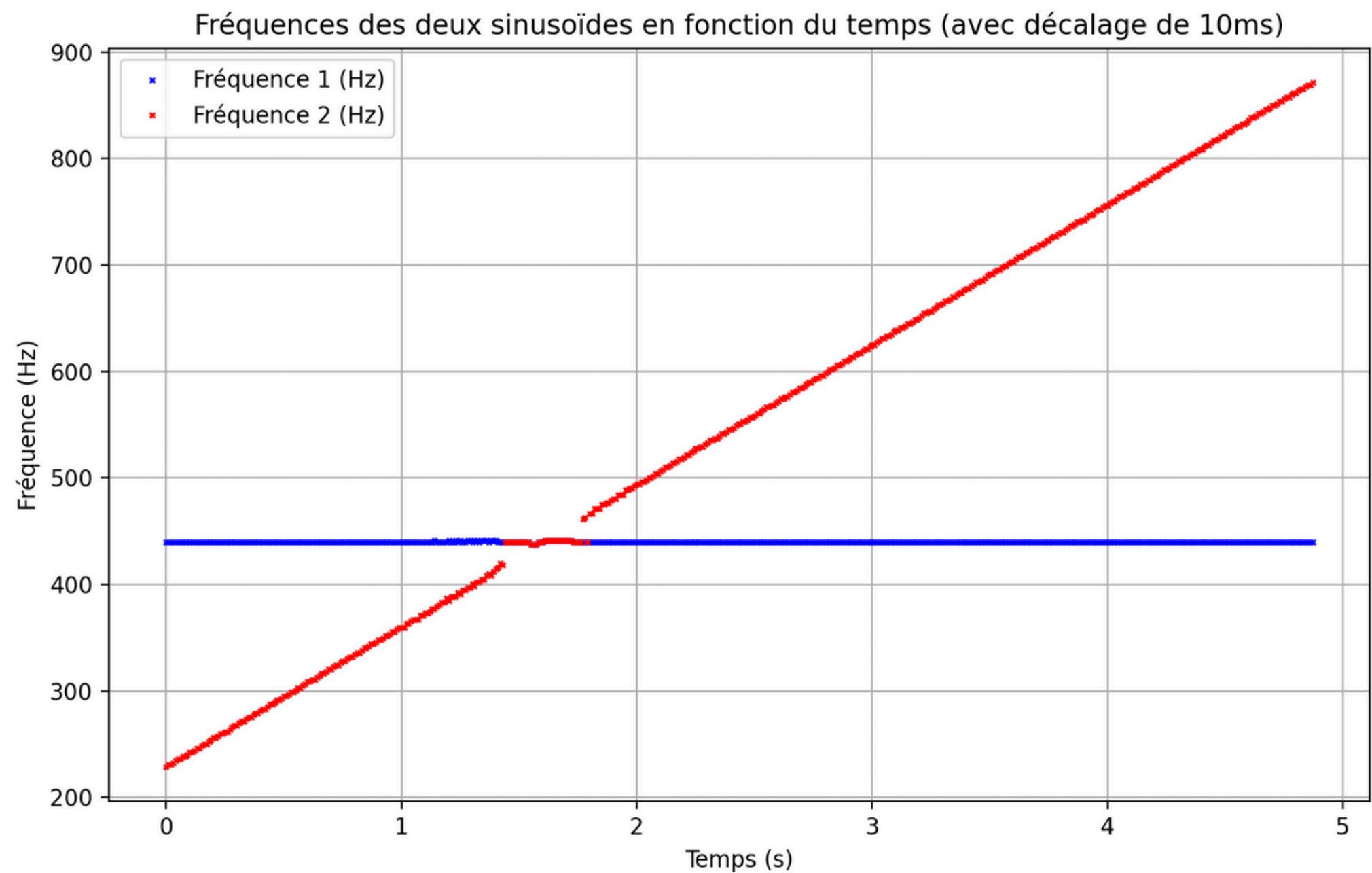
SEPARATION DES SINUS

Méthodes auto-régressives

Résultats :

Erreur supérieure aux méthodes testées précédemment

Trames avec erreur > 1% pour
Signal 2 :
29/493 (5.88%)



SEPARATION DES SINUS

Méthode MUSIC

Hypothèse mathématique :

$$x[n] = \sum_{k=1}^p A_k e^{j2\pi f_k n} + e[n]$$

Estimation de la matrice de covariance pour chaque trame

$$\mathbf{R} = \mathbb{E}[\mathbf{x}\mathbf{x}^H]$$

1. Sous espace signal
2. Sous espace bruit

Pour chaque f :

$$\mathbf{a}(f) = \begin{bmatrix} 1 \\ e^{-j2\pi f} \\ e^{-j4\pi f} \\ \vdots \\ e^{-j2\pi(p-1)f} \end{bmatrix}$$

vecteur de direction

$$P(f) = \frac{1}{\|\mathbf{S}_{\text{bruit}}^H \mathbf{a}(f)\|^2}$$

Pseudo spectre

$P(f)$ atteint un max =>
 f fréquence significative

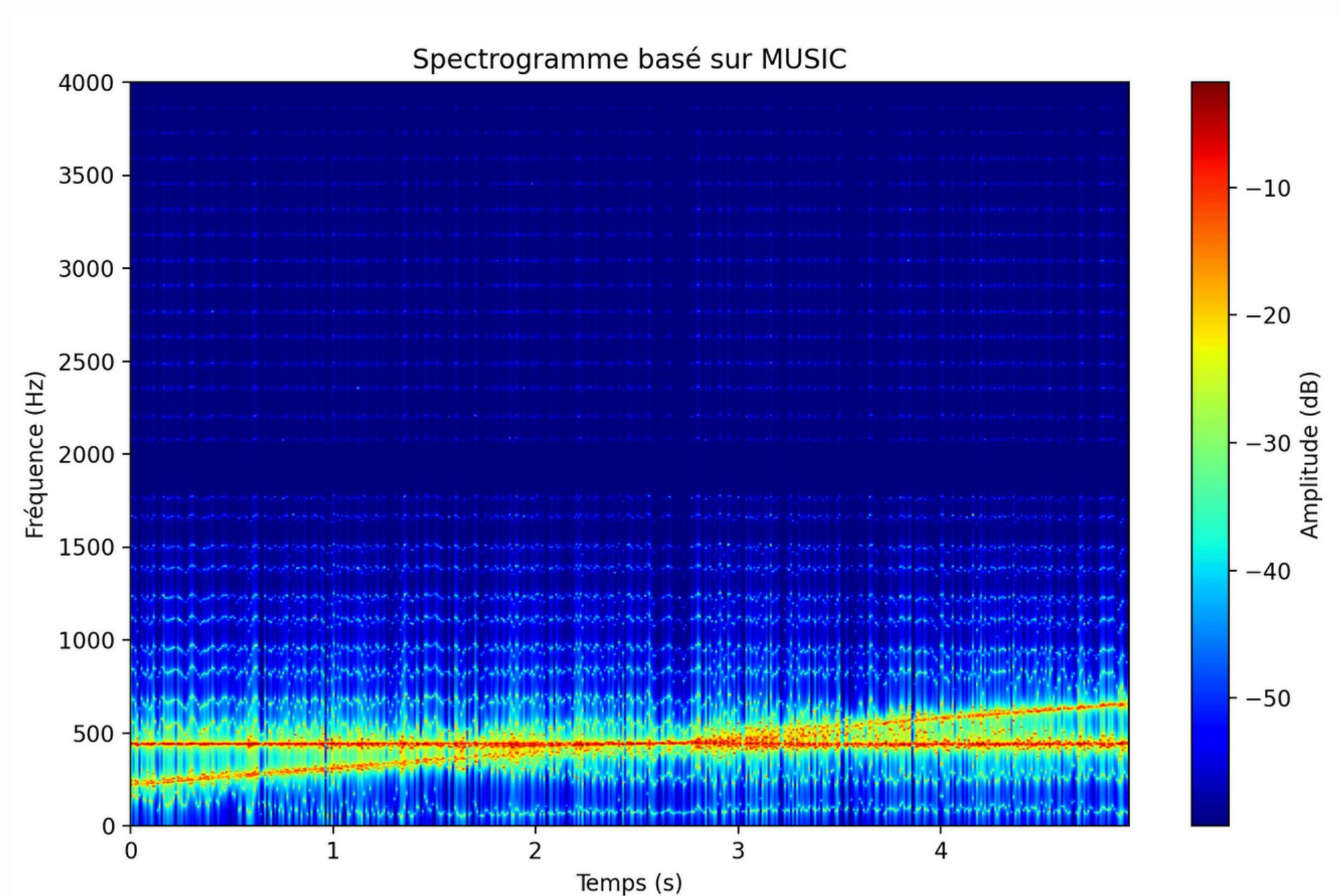
SEPARATION DES SINUS

Méthode MUSIC

Spectrogramme MUSIC :
croisement

Spectrogramme fortement bruité

séparation sous optimale au
croisement



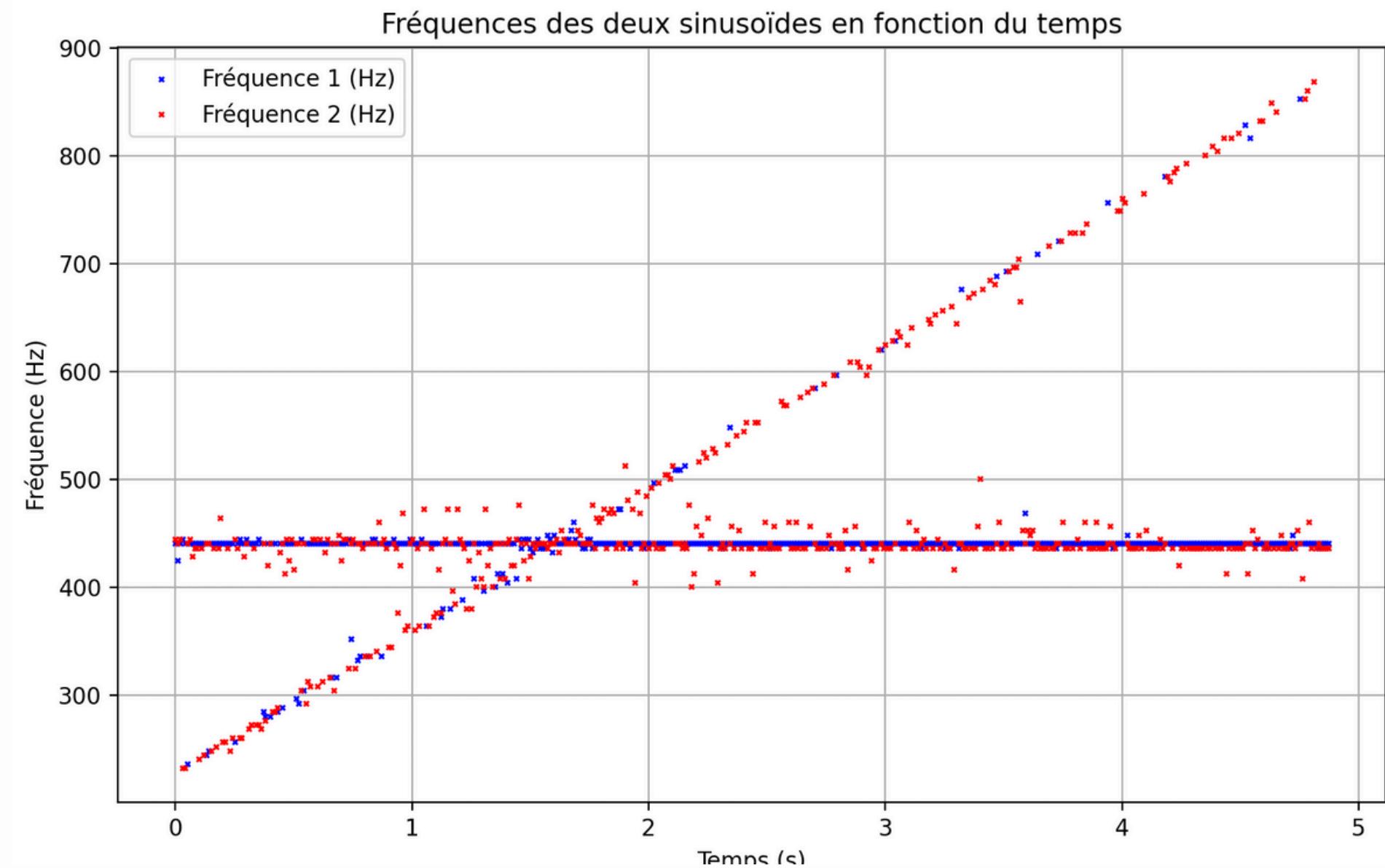
SEPARATION DES SINUS

Méthode MUSIC

Résultats décevants

Ici la méthode à été appliquée sans suivi des fréquences

on voit une très forte variabilité due au bruit même hors du croisement



SEPARATION DES SINUS

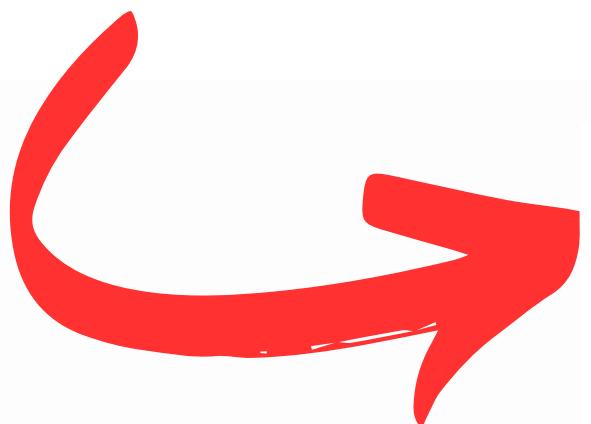
Méthodes auto-régressives

Méthode de Prony :

Modèle AR

Système linéaire donné par la corrélation croisée :

$$\sum_{m=0}^{N-p-1} x[m]x[m-k] = - \sum_{k=1}^p a_k \sum_{m=0}^{N-p-1} x[m]x[m-k]$$



$$T\mathbf{a} = -\mathbf{b}$$



$$\mathbf{a} = -T^{-1}\mathbf{b}$$

SEPARATION DES SINUS

Méthodes auto-régressives

Spectrogramme : croisement

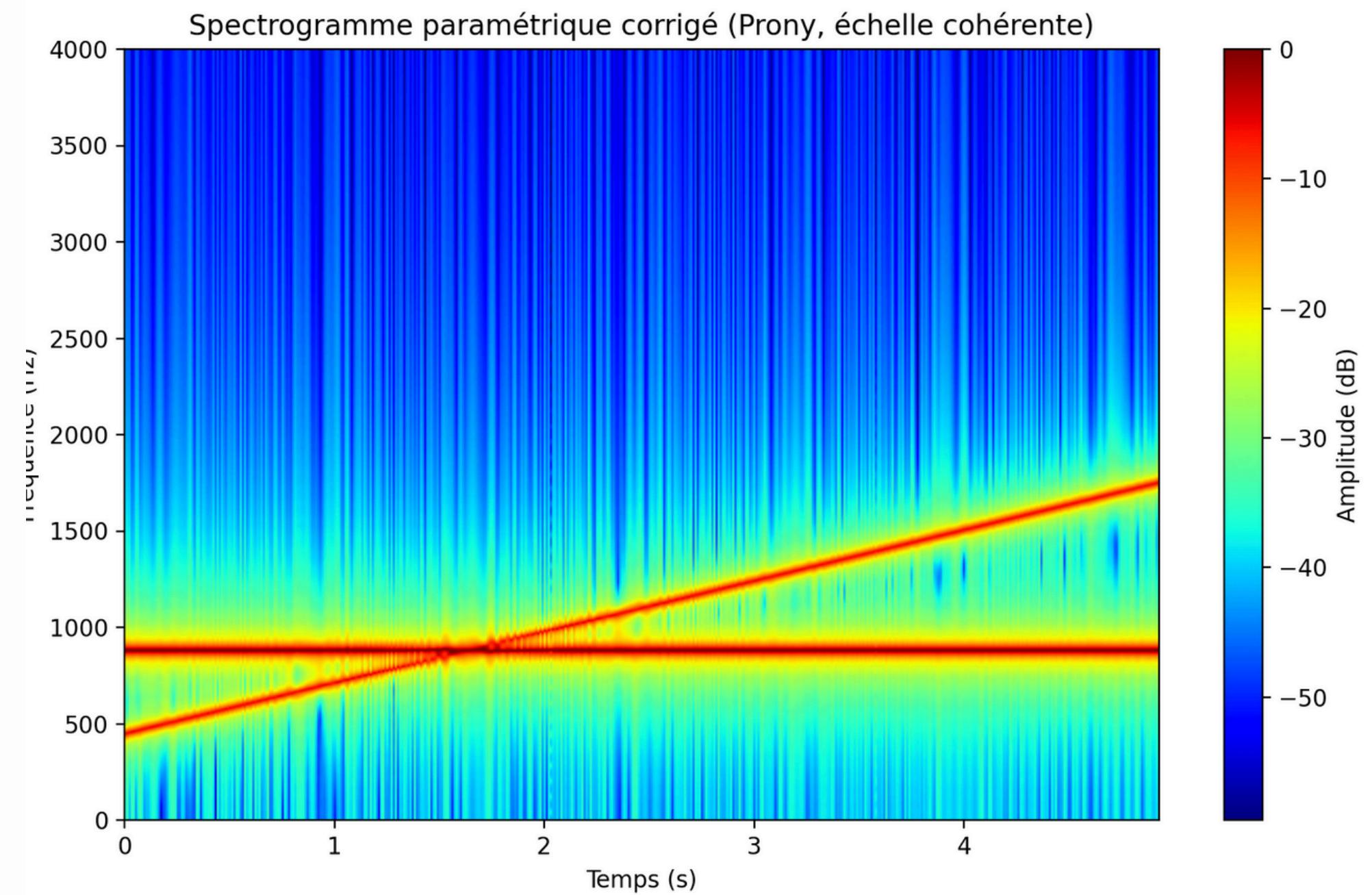
paramètres utilisés :

frame_size = $2^{**}9$

hop_size = 10ms

order = 50

n_fft = $2^{**}15$



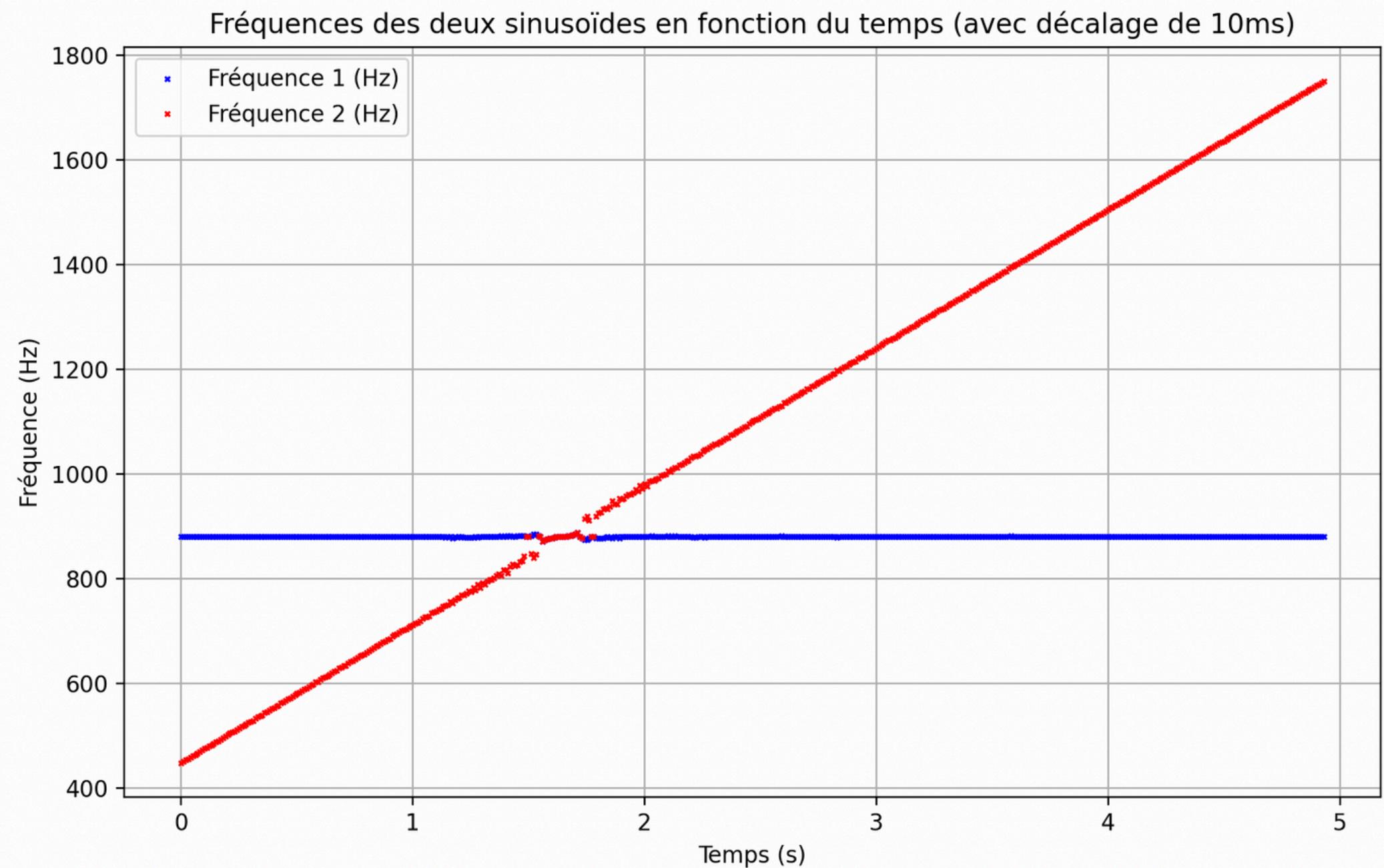
SEPARATION DES SINUS

Méthodes auto-régressives

Résultats :

Erreur moins élevée que pour
Burg

Trames avec erreur > 1% pour
Signal 2 :
19/493 (3.85%)

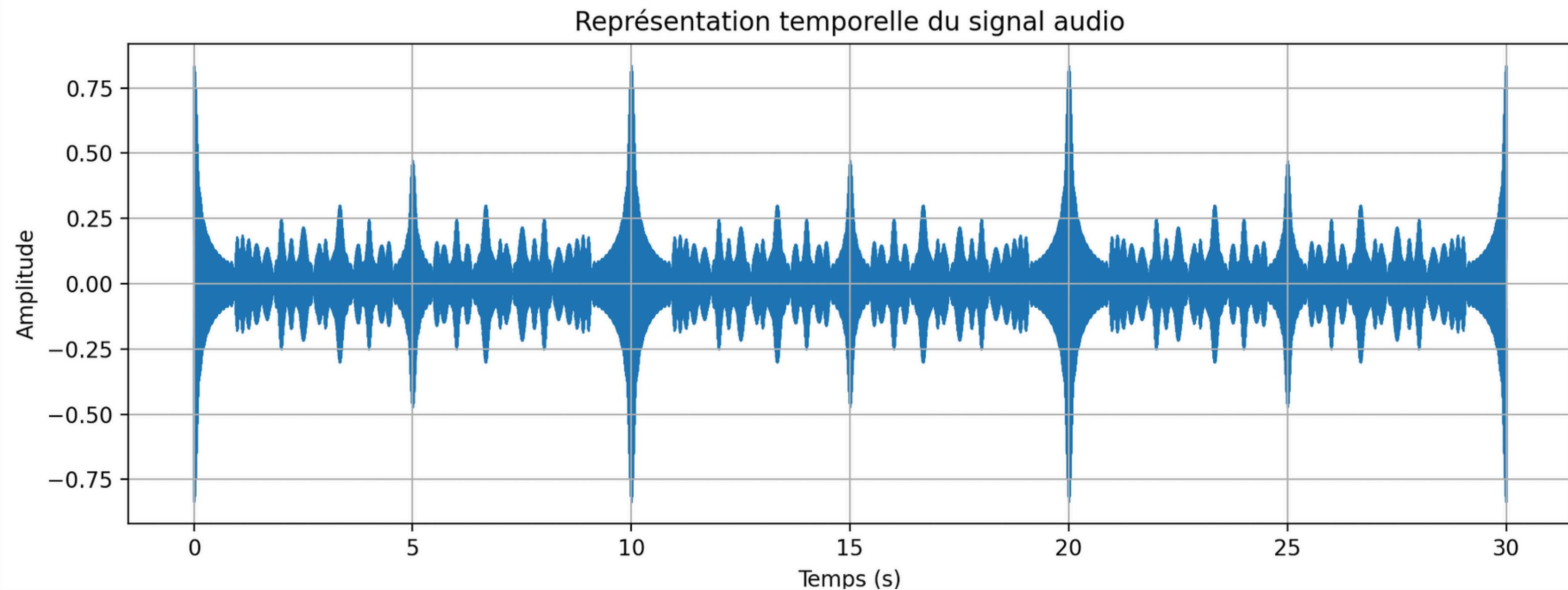


ANALYSE DE MYSON

ANALYSE MYSON.WAV

Traitement du signal

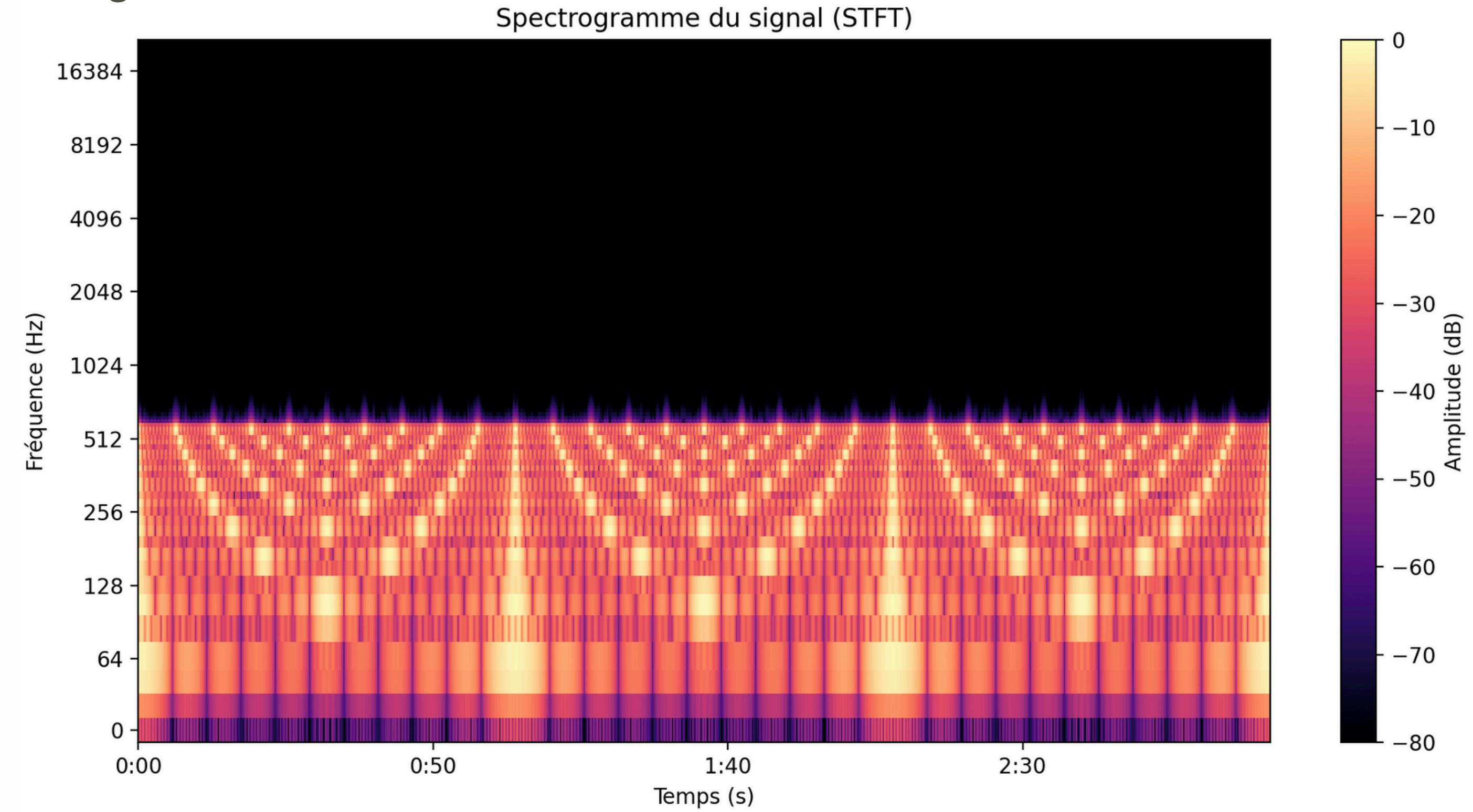
Analyse brute du signal :



ANALYSE MYSON.WAV

Traitement du signal

Spectrogramme :

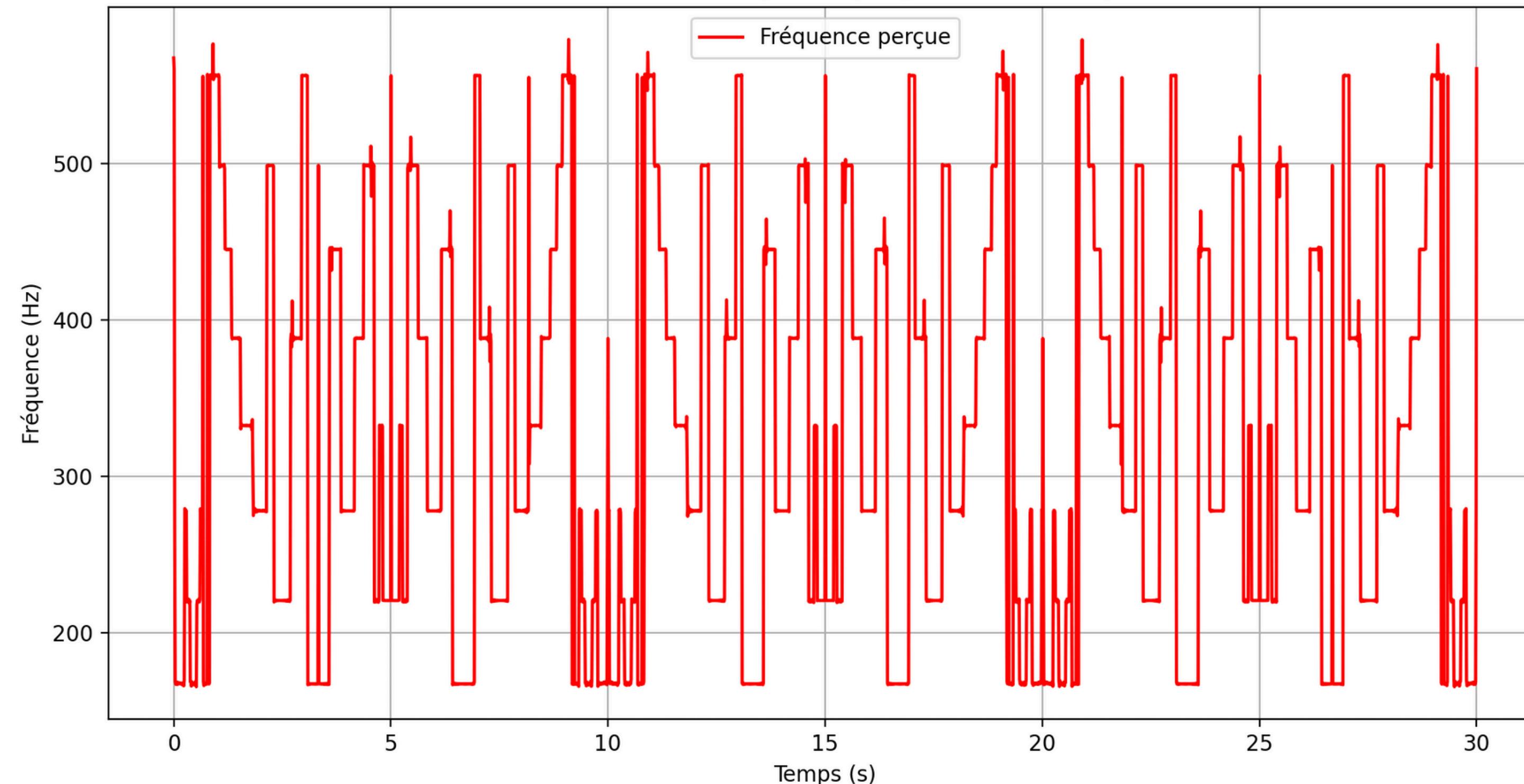


ANALYSE MYSON.WAV

Perception Humaine

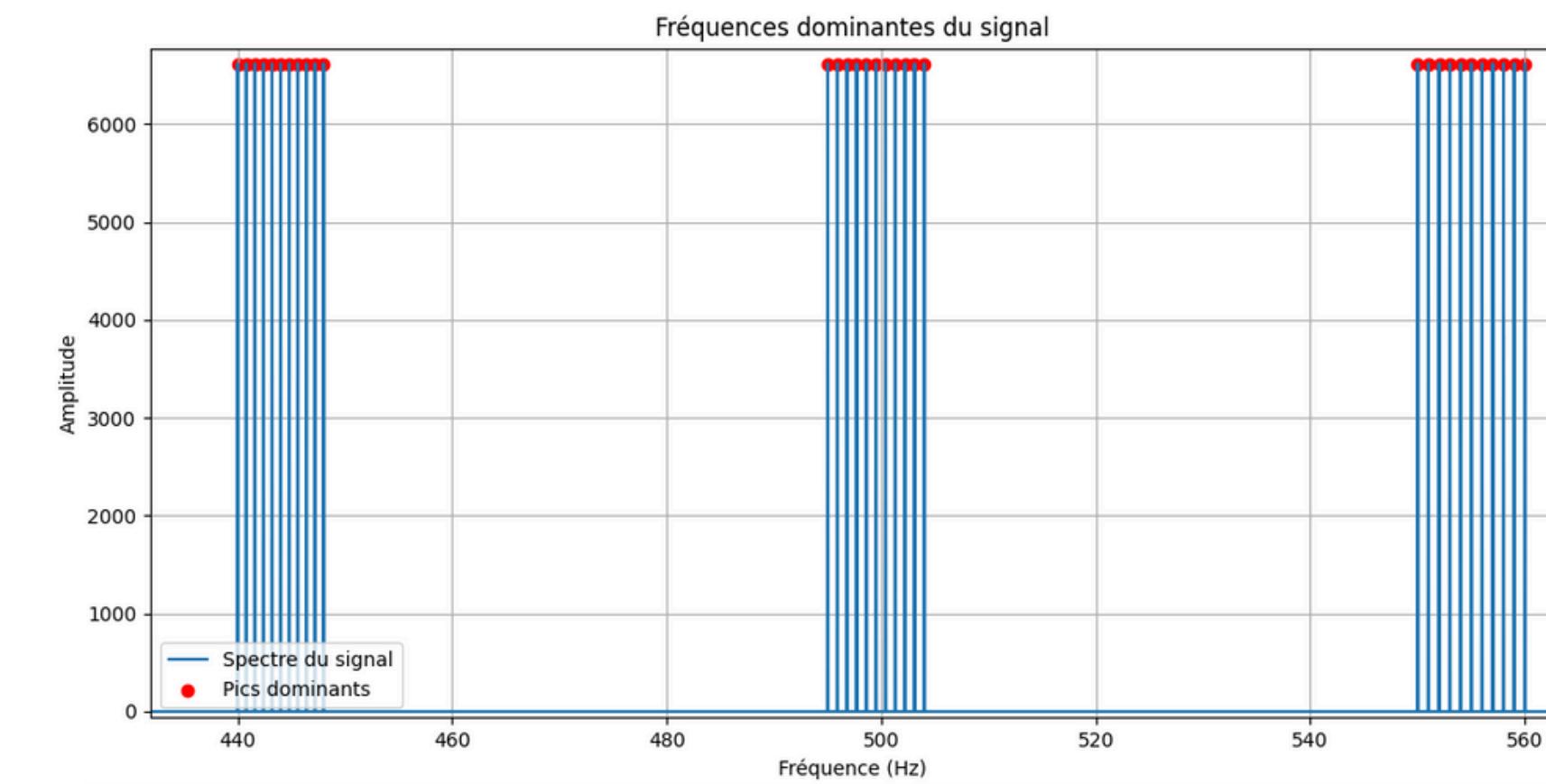
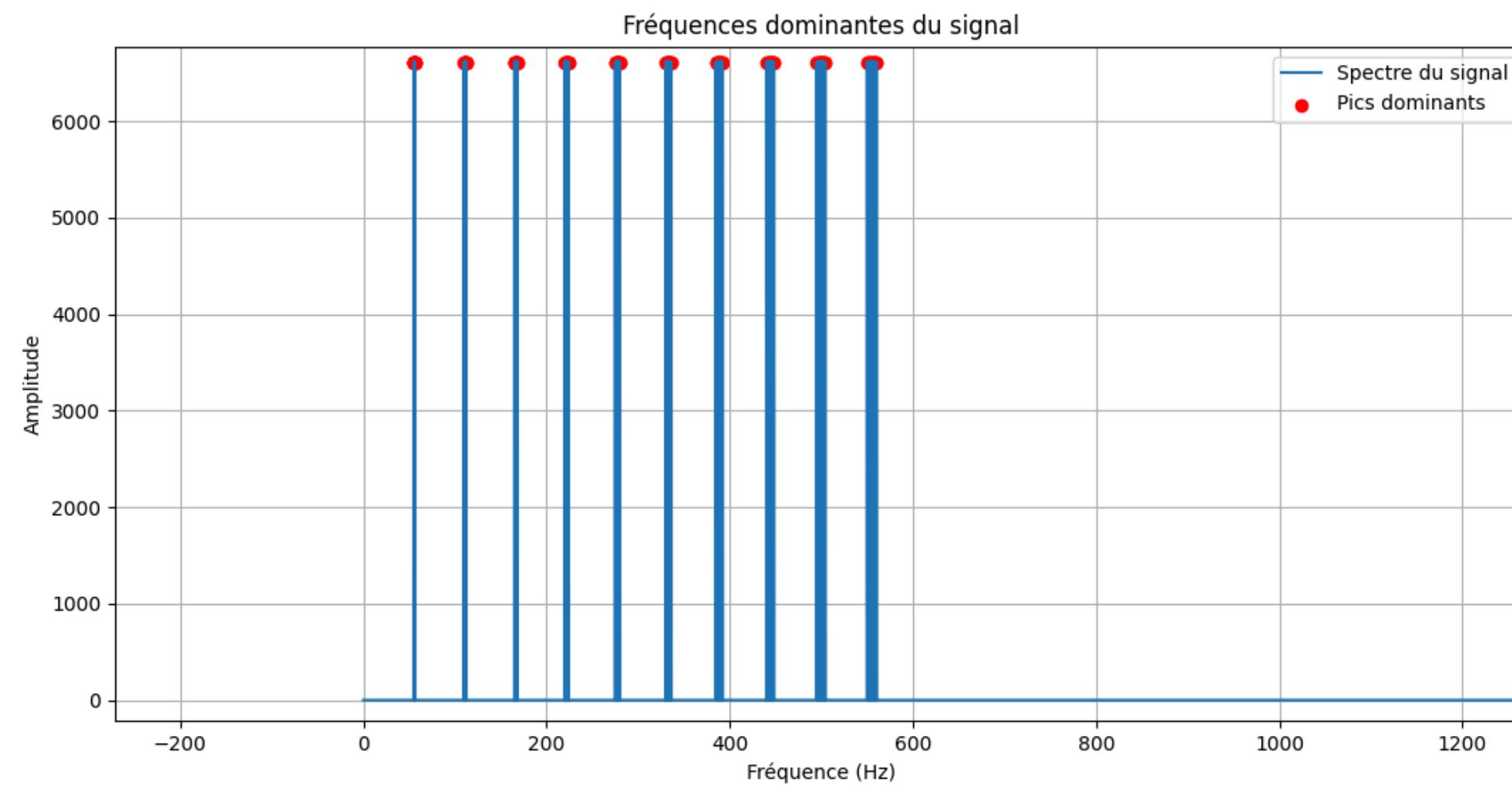
Fréquence perçue

Évolution de la fréquence perçue dans le temps



ANALYSE MYSON.WAV

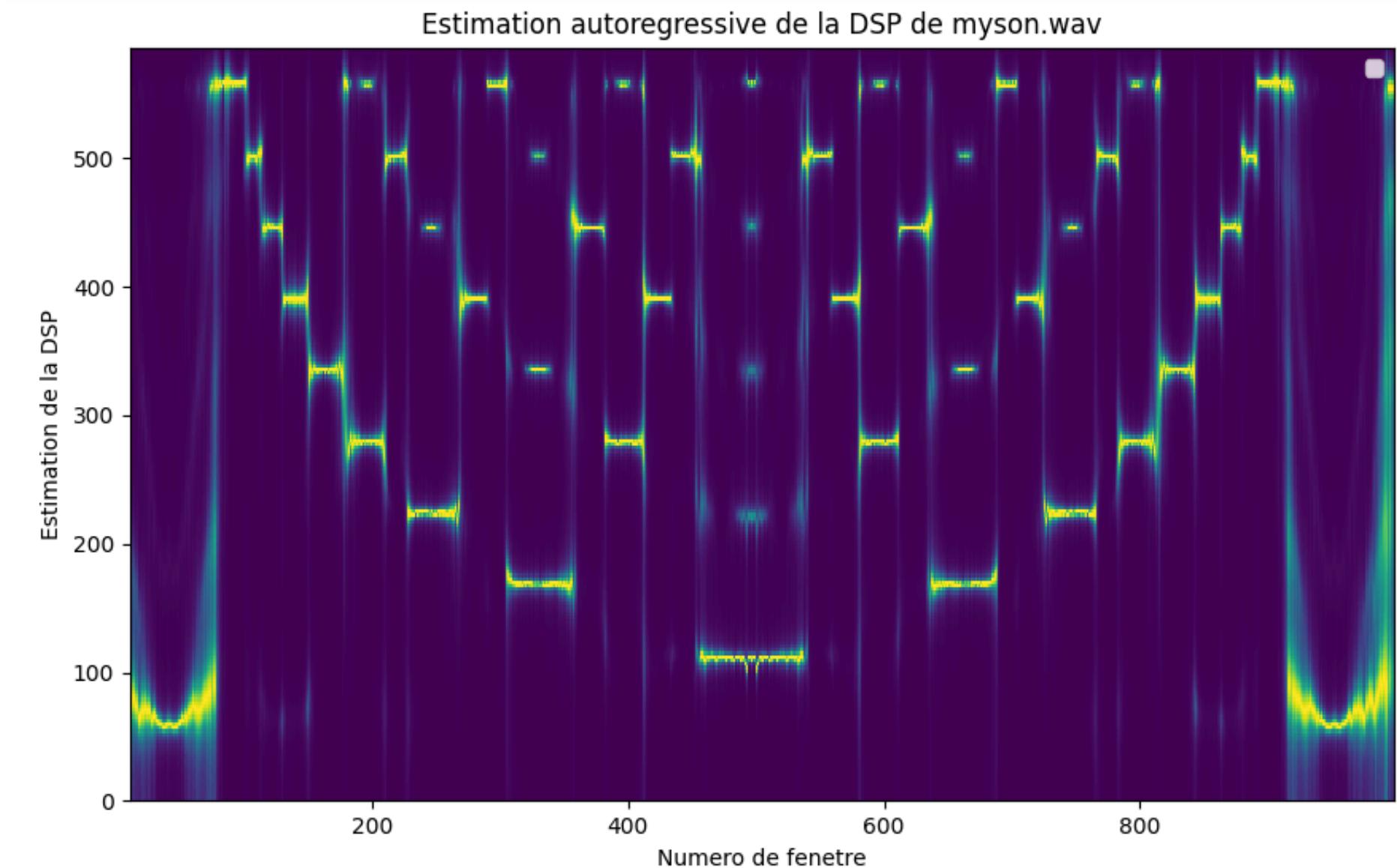
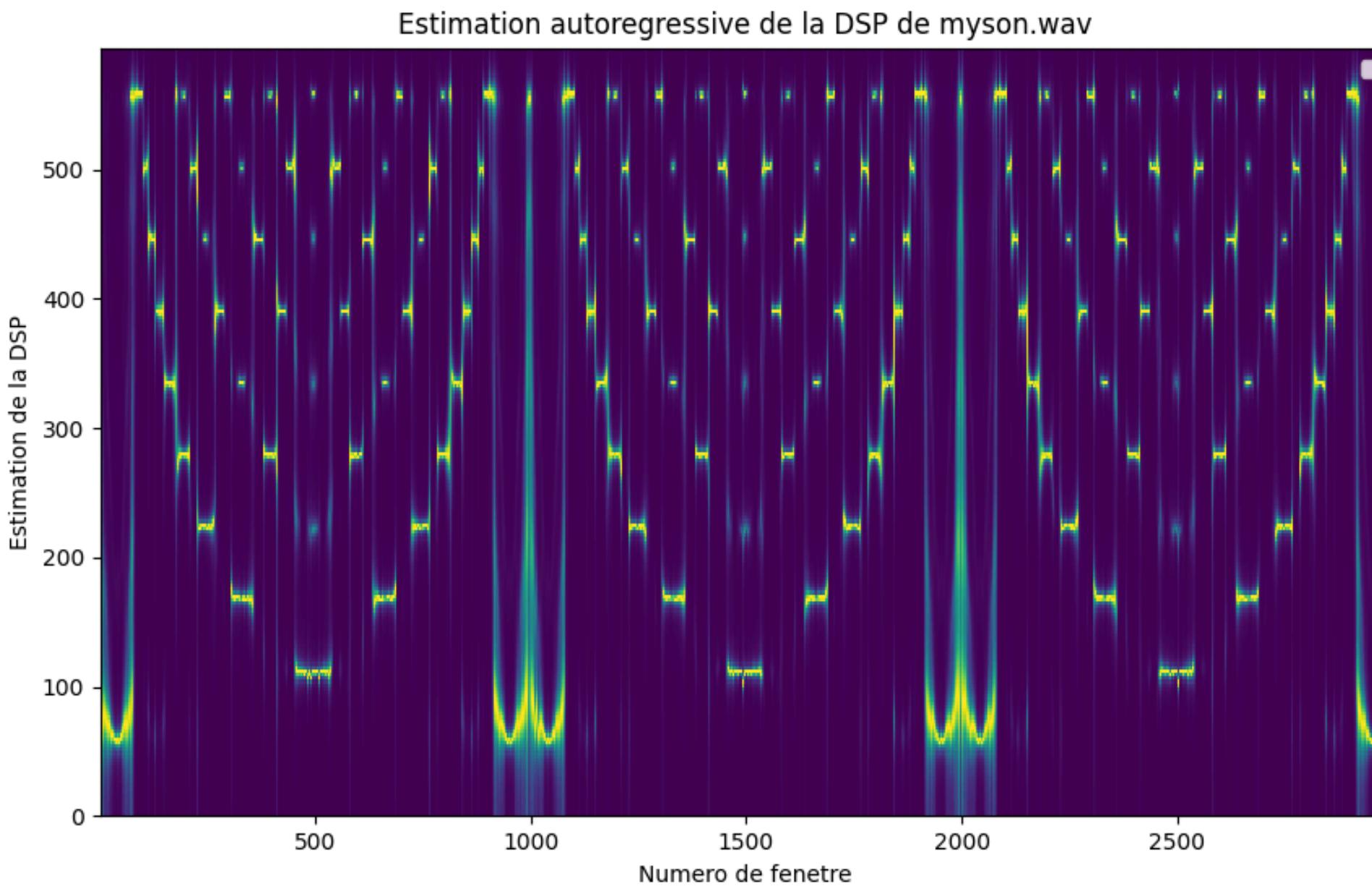
Perception Humaine
FFT de tout le signal



10 groupes de 11 pics qui s'épaissent

ANALYSE MYSON.WAV

Méthode de Levinson-Durbin



CONCLUSION