

TS224 – Filtrage et Estimation
Projet de traitement du signal de parole

Rehaussement de la parole par soustraction spectrale

Rédigé par :

Ben Amor Eya
Benzakri Meryem

Encadrant :

M. E. Grivel

Table des matières

1	Introduction	2
2	Préliminaire 1 : Analyse d'un bruit blanc gaussien	2
2.1	Objectif	2
2.2	Rappels théoriques	2
2.3	Étude de la fonction d'autocorrélation	2
2.4	Analyse fréquentielle et estimation spectrale	3
2.5	Platitude spectrale	3
2.6	Conclusion	4
3	Préliminaire 2 : Bruitage du signal de parole	4
3.1	Objectif	4
3.2	Ajout de bruit au signal de parole	4
3.3	Conclusion	5
4	Préliminaire 3 : Filtrage d'un signal de parole	5
4.1	Objectif	5
4.2	Caractérisation du filtre	5
4.3	Analyse expérimentale	6
4.4	Interprétation perceptive à l'écoute	6
4.5	Conclusion	7
5	Préliminaire 4 : Procédure d'addition–recouvrement (OLA)	7
5.1	Objectif	7
5.2	Principe de la procédure d'addition-recouvrement	7
5.3	Résultats de reconstruction	7
6	Soustraction spectrale et reconstruction du signal	8
6.1	Présentation de la méthode	8
6.2	Comparaison globale du signal rehaussé	9
6.3	Analyse des performances et limites de la méthode	9
6.4	Perception auditive	10
6.5	Conclusion	10
7	Conclusion générale	10

1 Introduction

La parole est un signal non stationnaire, complexe et riche en informations. Son traitement permet d'améliorer la qualité d'écoute, de réduire le bruit et de faciliter la reconnaissance automatique. Le projet TS224 – Filtrage et Estimation a pour objectif de mettre en œuvre un **rehaussement de la parole bruitée** par la méthode de **soustraction spectrale**, une approche classique et efficace pour la suppression de bruit additif.

2 Préliminaire 1 : Analyse d'un bruit blanc gaussien

2.1 Objectif

L'objectif de ce préliminaire est d'étudier expérimentalement les propriétés temporelles (autocorrélation) et fréquentielles (DSP) d'un bruit blanc gaussien (BBG) de moyenne nulle et de variance σ^2 .

2.2 Rappels théoriques

Un bruit blanc gaussien centré $\{b[n]\}$ est un processus aléatoire stationnaire de moyenne nulle, dont la fonction d'autocorrélation est une impulsion de variance σ^2 , centrée en 0. Sa densité spectrale de puissance (DSP) théorique est donc constante sur l'ensemble des fréquences : $S_b(f) = \sigma^2$.

2.3 Étude de la fonction d'autocorrélation

À partir d'une séquence finie, l'autocorrélation peut être estimée à l'aide d'un estimateur biaisé ou non biaisé.

- **Non biaisé** : $\hat{R}_{xx}(k) = \frac{1}{N-k} \sum_{n=k}^{N-1} x[n]x[n-k]$ → exact en moyenne, variance élevée aux grands décalages
- **Biaisé** : $\hat{R}_{xx}(k) = \frac{1}{N} \sum_{n=k}^{N-1} x[n]x[n-k]$ → stable, moins bruité et légère sous-estimation en amplitude

Observation expérimentale : La Fig 1 met en évidence les différences entre l'autocorrélation théorique d'un bruit blanc gaussien et ses estimations biaisée et non biaisée.

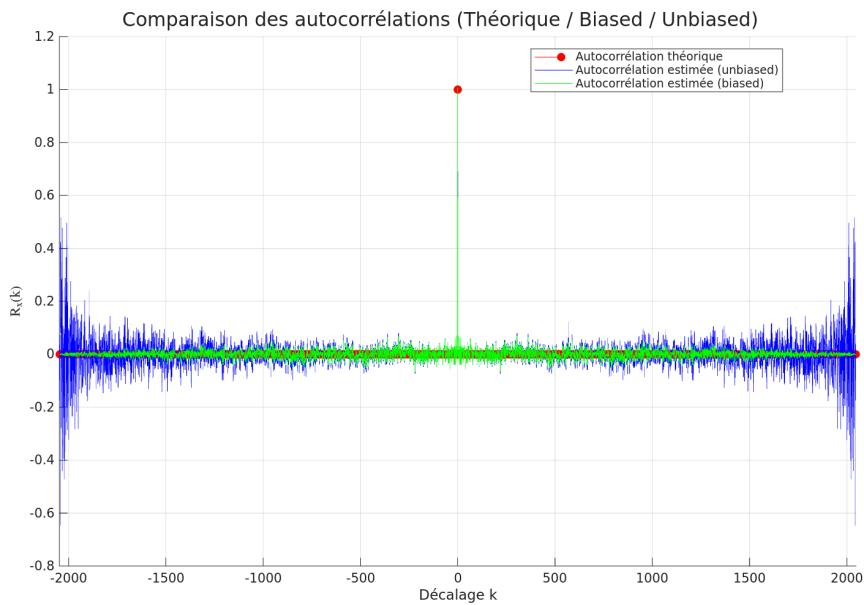


FIGURE 1 – Comparaison des autocorrélations théorique, biaisée et non biaisée d'un bruit blanc gaussien.

On observe que :

- l'autocorrélation théorique d'un bruit blanc gaussien est une impulsion de variance σ^2 en $k = 0$ et nulle ailleurs
- pour $k \neq 0$, les estimations fluctuent autour de zéro en raison du caractère statistique de l'estimation sur une séquence finie
- l'estimateur non biaisé devient fortement bruité pour les grands décalages, tandis que l'estimateur biaisé fournit une estimation plus stable, au prix d'une légère sous-estimation en amplitude

Ces observations confirment que le bruit blanc est un processus non corrélé pour tout décalage non nul, et que seule la valeur en $k = 0$ contient de l'énergie.

2.4 Analyse fréquentielle et estimation spectrale

L'étude fréquentielle du bruit blanc gaussien repose sur l'estimation de sa densité spectrale de puissance (DSP), théoriquement constante. Plusieurs méthodes d'estimation ont été mises en œuvre, chacune présentant un compromis entre variance, résolution fréquentielle et robustesse :

- **Périodogramme simple** : bonne résolution, variance élevée (spectre fluctuant)
- **Méthode de Daniell** : lissage fréquentiel réduisant la variance, perte de résolution
- **Méthode de Bartlett** : moyenne sur segments non recouvrants, variance réduite
- **Méthode de Welch** : segments recouvrants + fenêtrage Hamming, meilleur compromis
- **Corrélogramme** : FFT de l'autocorrélation (Wiener-Khintchine), sensible aux erreurs d'estimation

Comparaison expérimentale La Fig 2 illustre les estimations de DSP obtenues avec ces différentes méthodes pour un bruit blanc gaussien centré.

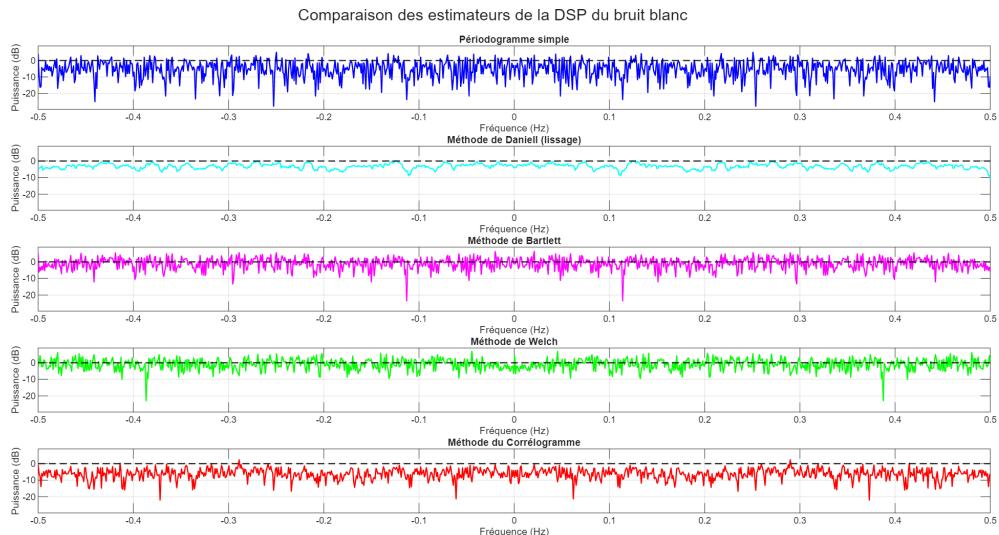


FIGURE 2 – Comparaison des estimateurs de la densité spectrale de puissance (DSP) du bruit blanc gaussien.

On remarque que la variance spectrale diminue sensiblement du périodogramme simple aux méthodes par moyennage (Bartlett, Welch). Pour un bruit blanc de variance σ^2 , elle passe d'environ σ^4 à un facteur réduit K , et diminue encore plus avec Welch grâce au recouvrement. La DSP reste globalement constante, les écarts reflétant uniquement la variance d'estimation.

2.5 Platitude spectrale

La *platitude spectrale* permet de quantifier le **degré de blancheur d'un signal**. Elle est définie comme le rapport entre la moyenne géométrique et la moyenne arithmétique des valeurs de la densité spectrale de puissance :

$$F = \frac{\left(\prod_{k=1}^N P[k] \right)^{1/N}}{\frac{1}{N} \sum_{k=1}^N P[k]}.$$

Résultats des réalisations : L'évaluation de cet indicateur a été réalisée sur 100 réalisations indépendantes du bruit blanc gaussien. On obtient :

- **Périodogramme** : $\bar{F}_{\text{périodogramme}} = 0.56325$, $\sigma_F = 0.0187$
- **Welch** : $\bar{F}_{\text{Welch}} = 0.80132$, $\sigma_F = 0.01129$.

La faible platitude du périodogramme ($\bar{F}_{\text{périodogramme}}$) s'explique par l'absence de moyennage temporel et de lissage fréquentiel, produisant un spectre très bruité. Welch (\bar{F}_{Welch}) améliore sensiblement la stabilité grâce au recouvrement et au fenêtrage Hamming, réduisant les fluctuations locales. L'écart à $F = 1$ provient du nombre fini d'échantillons et de la variance d'estimation.

Généralisation : La platitude spectrale peut être généralisée en remplaçant la moyenne géométrique par une *moyenne d'ordre p* :

$$M_p(P) = \left(\frac{1}{N} \sum_{k=1}^N P[k]^p \right)^{1/p}, \quad F_p = \frac{M_p(P)}{M_1(P)}.$$

Cette formulation permet d'ajuster la sensibilité de la mesure à la répartition des puissances spectrales :

- $p \rightarrow 0$: platitude classique (moyenne géométrique)
- $0 < p < 1$: tolérante aux variations locales
- $p > 1$: stricte, sensible aux pics spectraux

Donc, plus p augmente, plus l'indicateur pénalise les spectres irréguliers.

2.6 Conclusion

Cette première étape du projet a permis de vérifier expérimentalement les propriétés fondamentales du bruit blanc gaussien : $R(k \neq 0) \approx 0$, DSP constante, $F \approx 1$. Les méthodes par moyennage réduisent la variance au prix de la résolution. Welch offre le meilleur compromis et sera privilégiée.

3 Préliminaire 2 : Bruitage du signal de parole

3.1 Objectif

Le but de cette section est d'étudier l'effet d'un bruit blanc additif sur un signal de parole pour différents niveaux de RSB (5, 10, 15 dB) en analysant l'impact du bruit sur les structures temporelles et spectrales du signal.

3.2 Ajout de bruit au signal de parole

Un bruit blanc gaussien avec $\sigma_b = \sqrt{P_{\text{signal}}/10^{\text{RSB}_{\text{dB}}/10}}$ est ajouté au signal de parole selon un RSB cible $b[n] = \sqrt{P_{\text{bruit}}} \times \text{randn}(n)$. La fonction `ajouter_bruit` a été développée pour introduire ce bruit au signal de parole. Le RSB (dB) contrôle le niveau de dégradation : plus il est faible, plus le bruit est dominant.

Résultats et analyse

Le signal de parole (`fcno01fz`) a été échantillonné à $F_e = 8$ kHz et bruité pour plusieurs valeurs de RSB. La Fig. 3 illustre le signal de parole original et les signaux bruités correspondants :

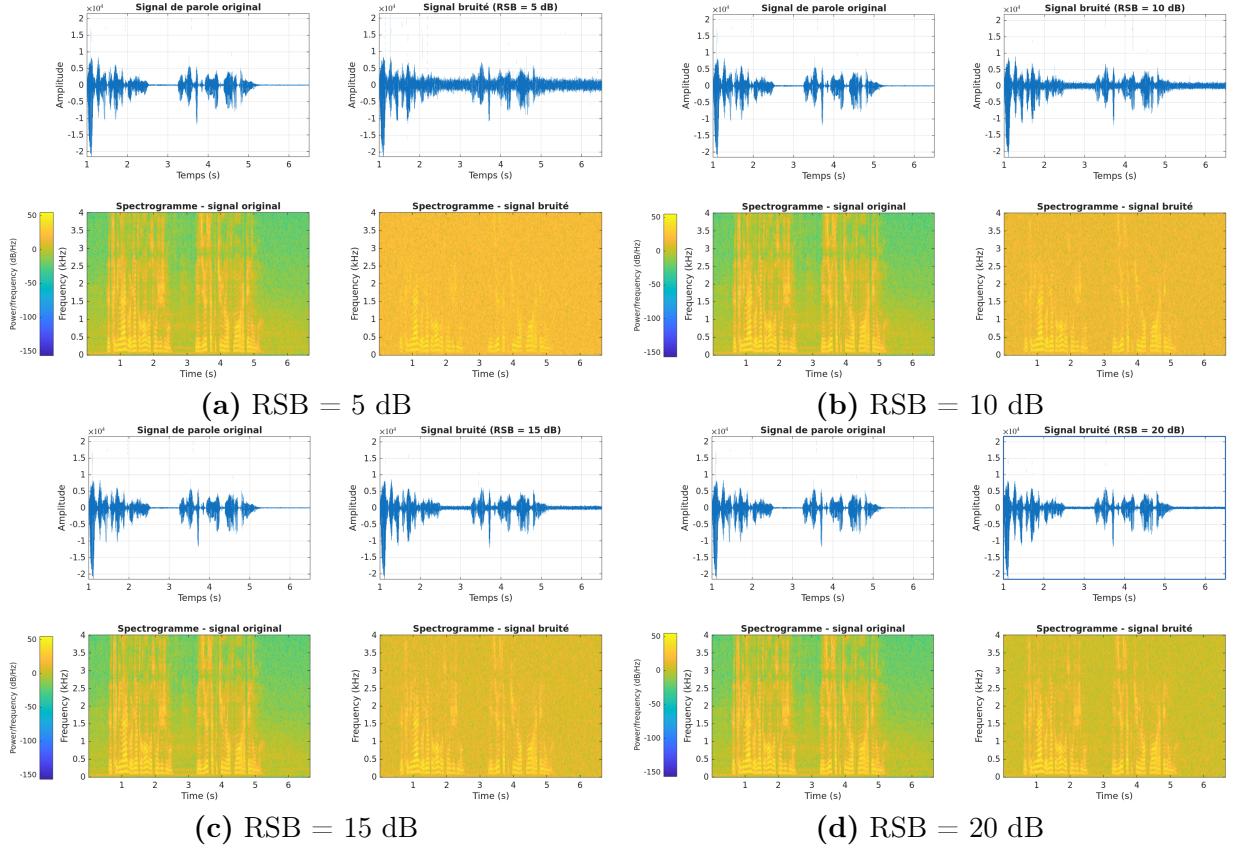


FIGURE 3 – Évolution du signal de parole bruité pour différents rapports signal/bruit (RSB).

Analyse temporelle : Amplitude instantanée augmentée, zones de silence « remplies » par le bruit. La forme originale devient moins reconnaissable pour les faibles RSB.

Analyse fréquentielle : Les spectrogrammes confirment ces observations :

- **5 dB** : formants masqués, spectre quasi-uniforme (bruit dominant)
- **10 - 15 dB** : composantes harmoniques partiellement visibles
- **20 dB** : parole claire, formants structurés et identifiables

3.3 Conclusion

L'ajout de bruit blanc simule des conditions de transmission dégradées. Un RSB faible masque la parole visuellement et perceptivement. L'augmentation du RSB améliore la structure harmonique et l'intelligibilité.

4 Préliminaire 3 : Filtrage d'un signal de parole

4.1 Objectif

L'objectif de ce préliminaire est d'étudier les effets d'un filtre défini par la réponse impulsionnelle : $h[k] = \delta[k] + \delta[k - k_0]$, pour différentes valeurs de décalage $k_0 = 1, 10, 100, 300$. Ce filtre simple permet d'illustrer l'impact du décalage temporel sur le spectre du signal et sur la perception du son.

4.2 Caractérisation du filtre

Nature du filtre : Filtre RIF (Réponse Impulsionnelle Finie), car 2 échantillons non nuls, causal et stable : $\sum_k |h[k]| = 2 < \infty$.

Fonction de transfert : La transformée en z de $h[k]$ est : $H(z) = 1 + z^{-k_0}$.

Pôles et zéros : Zéros uniformément répartis sur le cercle unité : $z_m = e^{j\pi(2m+1)/k_0}$ avec $m = 0, 1, \dots, k_0 - 1$, d'où la structure périodique du spectre.

Réponse en fréquence : En remplaçant $z = e^{j2\pi f/F_e}$, on obtient : $H(f) = 2 \cos(\pi f k_0 / F_e) e^{-j\pi f k_0 / F_e}$. Le module de la réponse fréquentielle est donc : $|H(f)| = 2 |\cos(\pi f k_0 / F_e)|$. On observe une structure périodique en fréquence, typique des *filtres à peigne*.

Phase et retard de groupe : La phase du filtre est : $\phi(f) = -\pi f k_0 / F_e$, et le retard de groupe correspondant est constant : $\tau_g = -\frac{1}{2\pi} \frac{d\phi}{df} = \frac{k_0}{2F_e}$. Le filtre se comporte donc comme un **filtre à phase linéaire** introduisant un léger retard moyen dépendant de k_0 .

4.3 Analyse expérimentale

La Fig. 4 présente la représentation réelle et le spectrogramme du signal de parole filtré pour différentes valeurs de k_0 .

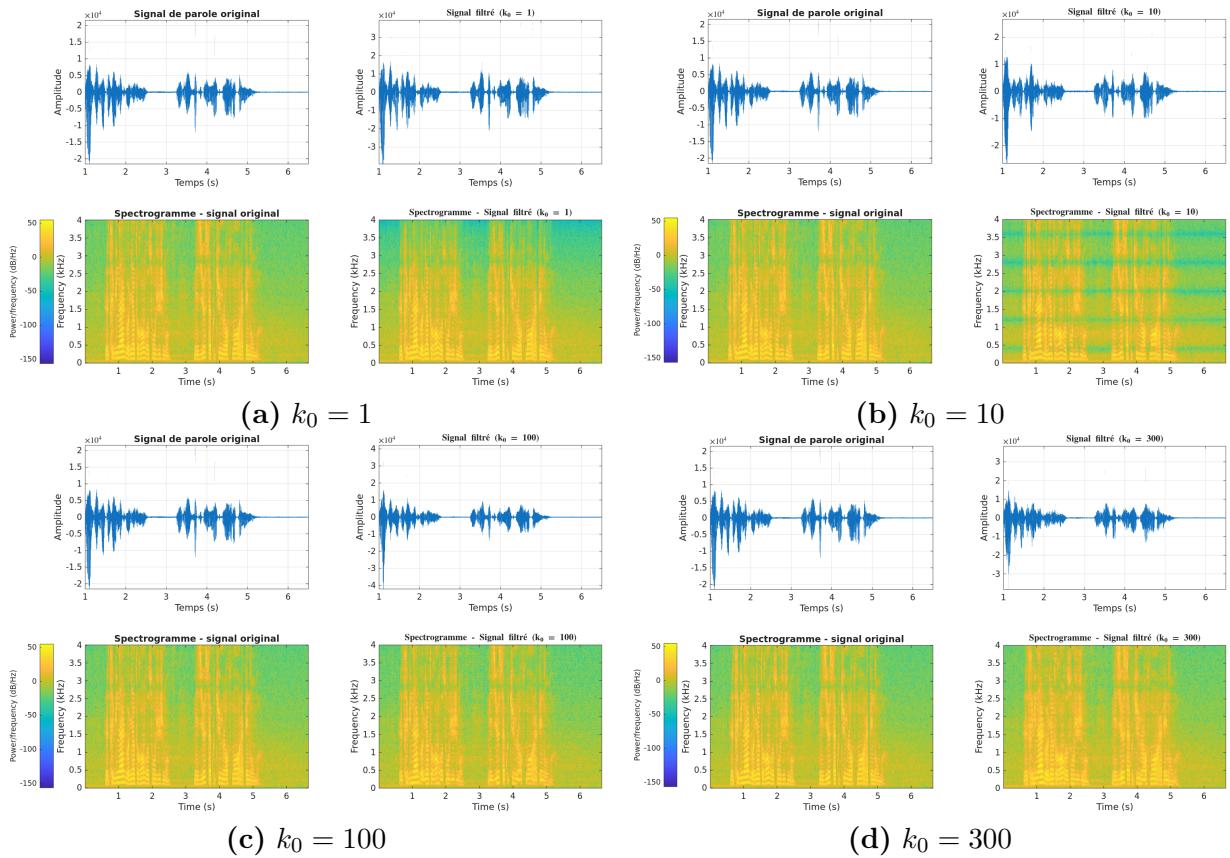


FIGURE 4 – Représentation réelle et spectrogramme du signal filtré pour différentes valeurs de k_0 .

Observations :

- $k_0 = 1$: Amplification simple, effet imperceptible
- $k_0 = 10$: Oscillations régulières, périodicité F_e/k_0
- $k_0 = 100$: Bandes horizontales fréquentes et plus rapprochées, traduisant la **structure en peigne**
- $k_0 = 300$: Spectrogramme dense, motifs quasi continus

Conclusion : Ces résultats confirment qu'un long retard (k_0 grand) génère un peigne aux dents étroites et rapprochées, traduisant la périodicité F_e/k_0 imposée par $H(f) = 1 + e^{-j2\pi f k_0 / F_e}$.

4.4 Interprétation perceptive à l'écoute

Afin de relier l'analyse théorique aux effets perceptibles, le signal de parole filtré a été écouté pour différentes valeurs de k_0 . La lecture a été réalisée à l'aide de la fonction `sound(y / max(abs(y)), Fe)` ;

de MATLAB où y est le signal filtré et F_e la fréquence d'échantillonnage. Cette normalisation évite toute saturation audio.

Observations auditives : À l'écoute des signaux filtrés pour différentes valeurs de k_0 , on constate une évolution marquée de la coloration sonore :

- Pour $k_0 = 1$: Timbre inchangé (amplification simple)
- Pour $k_0 = 10$: Léger effet flanger
- Pour $k_0 = 100$: Résonance métallique/écho court
- Pour $k_0 = 300$: Réverbération prononcée, son « doublé »

Ces observations confirment que l'augmentation du retard resserre les dents du peigne, rendant les interférences fréquentielles audibles.

4.5 Conclusion

Ce préliminaire illustre comment un simple filtre à deux impulsions engendre un effet de peigne fréquentiel marqué, démontrant la relation directe entre retard temporel k_0 et perception auditive.

5 Préliminaire 4 : Procédure d'addition–recouvrement (OLA)

5.1 Objectif

Ce préliminaire vise à mettre en place une procédure d'**addition–recouvrement** permettant de traiter un signal non stationnaire par trames successives tout en assurant une reconstruction fidèle du signal global. Cette approche est essentielle dans le traitement des signaux de parole bruités, notamment lorsque le bruit est aléatoire ou non stationnaire.

5.2 Principe de la procédure d'addition–recouvrement

- **Découpage** : Signal $x[n]$ divisé en trames $x_m[n]$ de longueur L avec recouvrement de 50% (pas de trame $R = L/2$)
- **Fenêtrage** : Chaque trame $x_m[n]$ multipliée par une fenêtre de Hamming $w[n]$ pour limiter les effets de bord
- **Reconstruction** : Sommation pondérée

$$x_{\text{recon}}[n] = \frac{\sum_m x_m[n]w[n - mR]}{\sum_m w[n - mR]}$$

- **Normalisation** : Le dénominateur (poids) $\sum_m w[n - mR]$ compense les recouvrements, garantissant $\sum_m w[n - mR] = \text{constante}$ pour une reconstruction exacte

5.3 Résultats de reconstruction

Pour évaluer la fidélité de la procédure d'addition–recouvrement (OLA), nous appliquons la méthode à un signal synthétique : $x(t) = \sin(2\pi 50t) + 0.5 \sin(2\pi 120t)$, échantillonné à $f_s = 1600$ Hz et découpé en trames de longueur $L = 32$ avec un pas $R = L/8$, c'est-à-dire un recouvrement très important, en utilisant une fenêtre de Hann.

Choix de la fenêtre : La fenêtre de Hann, combinée à un recouvrement important, garantit que la somme des fenêtres reste constante ($\sum_m w[n - mR] = \text{constante}$), assurant une reconstruction exacte tout en éliminant les discontinuités entre trames grâce à ses bords nuls.

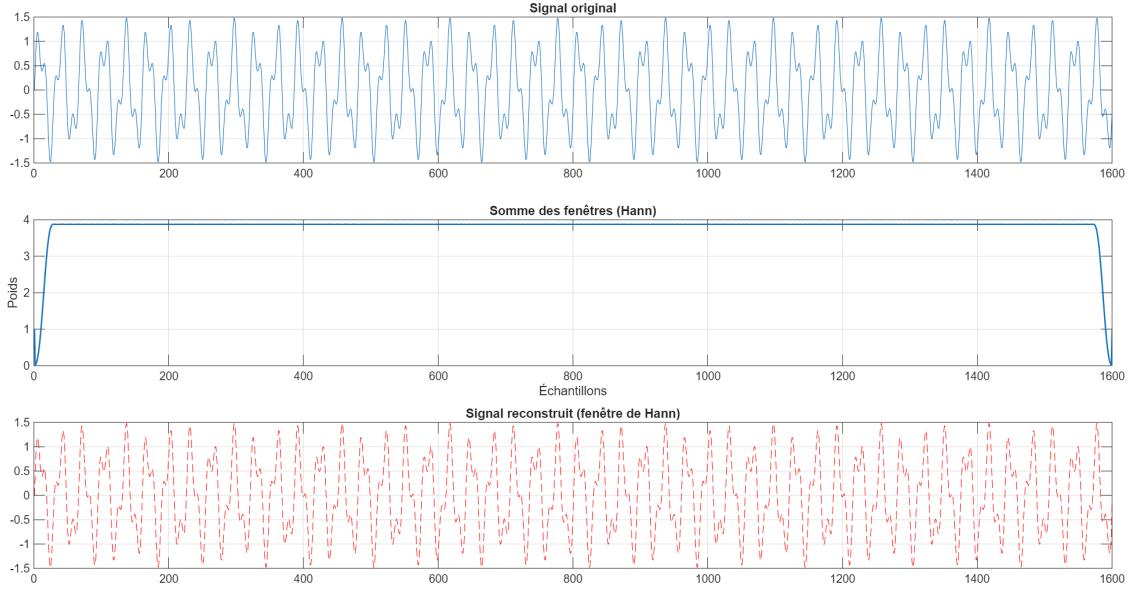


FIGURE 5 – Signal original, poids de recouvrement et signal reconstruit (fenêtre de Hann).

La Fig 5 montre les résultats de la procédure OLA. La somme des fenêtres reste lisse et stable, ce qui garantit une addition cohérente des trames et une reconstruction fiable.

Erreur globale :

$$\text{EQM}_{\text{Hann}} = 1.11 \times 10^{-4}, \quad \max |e_{\text{Hann}}[n]| = 4.22 \times 10^{-1}$$

Cette erreur est principalement concentrée aux bords du signal, où le recouvrement est incomplet. On a une erreur d'ordre 10^{-16} au centre du signal, car la fenêtre de Hann assure un recouvrement très régulier sur cette zone.

6 Soustraction spectrale et reconstruction du signal

6.1 Présentation de la méthode

- On observe le signal : $y[n] = s[n] + b[n]$ (parole + bruit blanc)
- On traite trame par trame ce signal, en appliquant la transformée de Fourier à court terme (STFT) : $Y_k(f) = S_k(f) + B_k(f)$
- Comme le signal de parole et le bruit sont supposés décorrélés, leurs densités spectrales de puissance vérifient : $|Y_k(f)|^2 = |S_k(f)|^2 + |B_k(f)|^2$

Étapes de traitement :

1. **Estimation du bruit** : $\hat{P}_b(f) = \frac{1}{M_b} \sum_{m=1}^{M_b} |B_m(f)|^2$ sur les 20 premières trames
2. **Soustraction spectrale** : $\hat{P}_s(f) = |Y(f)|^2 - \hat{P}_b(f)$, puis $\hat{P}_s(f) = \max(\hat{P}_s(f), 0)$ (half-wave rectification)
3. **Reconstruction amplitude** : $|\hat{S}(f)| = \sqrt{\hat{P}_s(f)}$
4. **Conservation de phase** : $\hat{S}(f) = |\hat{S}(f)| e^{j\phi_y(f)}$
5. **Reconstruction temporelle** : $\hat{s}_k[n] = \text{Re}\{\text{IFFT}(\hat{S}_k(f))\}$, puis OLA avec fenêtre de Hann

Analyse trame par trame

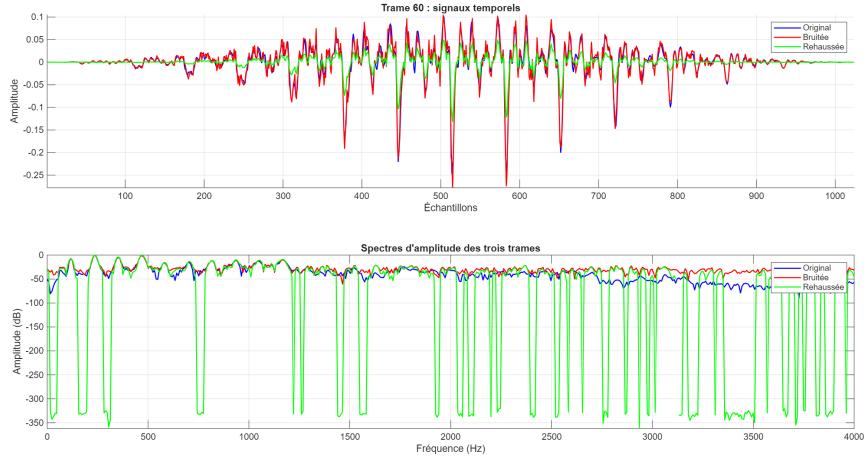


FIGURE 6 – Signaux temporels et spectres de la trame 60 : original, bruité, rehaussé.

La Fig. 6 (trame 60) montre le signal original, bruité et rehaussé. Le spectre rehaussé présente une réduction partielle du bruit mais avec des annulations brutales de composantes fréquentielles, caractéristiques du **noise music**.

6.2 Comparaison globale du signal rehaussé

La **reconstruction globale** a été réalisée par méthode OLA. La Fig. 7 montre les signaux temporels et leurs spectrogrammes.

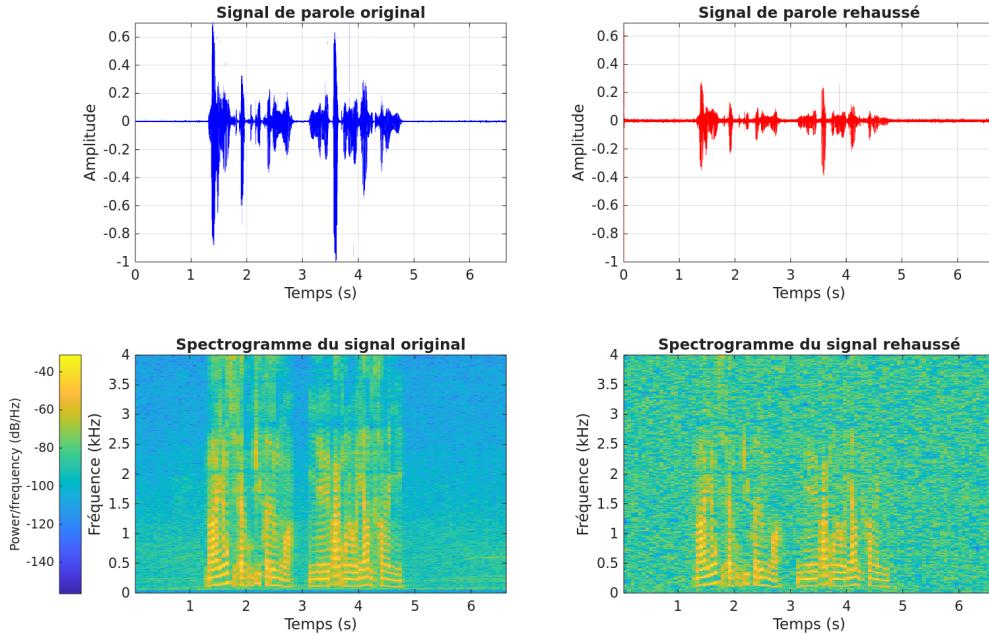


FIGURE 7 – Comparaison globale des signaux original et rehaussé.

Bien que certaines bandes de bruit soient atténuées, la parole est elle aussi affaiblie, et des artefacts non stationnaires apparaissent dans les zones silencieuses.

6.3 Analyse des performances et limites de la méthode

Mesures :

Les mesures obtenues pour une réalisation du bruit montrent que :

$$RSB_{\text{avant}} = 9.99 \text{ dB}, \quad RSB_{\text{aprs}} = 1.83 \text{ dB}, \quad \text{Gain RSB} = -8.16 \text{ dB}.$$

Évaluation statistique sur plusieurs niveaux de bruit :

Pour obtenir une analyse plus représentative, l'expérience a été répétée sur 20 réalisations indépendantes, pour trois niveaux de RSB initial : $RSB_{avant} = 5, 10, 15$ dB.

Les moyennes obtenues sont :

RSB avant (dB)	RSB après (dB)	Gain RSB (dB)
5	3.20	-2.41
10	4.80	-5.44
15	5.38	-9.47

TABLE 1 – Résultats moyens sur 20 réalisations indépendantes.

Observation : Dégradation systématique du RSB, plus marquée pour les signaux initialement propres.

Interprétation : Quand $|Y(f)|^2 \leq \hat{P}_b(f)$, la puissance est mise à zéro (half-wave rectification). Ces annulations instables d'une trame à l'autre créent des composantes spectrales fines, aléatoires et hautement perceptibles, ce qui engendre un artefact bien connu : **le noise music**.

Limites :

- Atténuation des zones faibles de la parole (confondues avec le bruit)
- Génération de noise music
- Dégradation globale du RSB

6.4 Perception auditive

Le signal rehaussé présente un bruit résiduel non stationnaire (granulaire, scintillant, métallique), très différent du bruit blanc initial. Il s'agit du **noise music**, artefact typique des méthodes de soustraction spectrale non régularisées.

6.5 Conclusion

La soustraction spectrale naïve (Boll, 1979) réduit le bruit mais introduit des artefacts audibles et dégrade le RSB. Des variantes régularisées sont nécessaires (sur-soustraction, spectral flooring, MMSE-STSA, Wiener adaptatif).

7 Conclusion générale

Ce projet a mis en œuvre les principales étapes du traitement de la parole : génération de bruit blanc gaussien, filtrage et effets perceptifs, validation de la procédure d'addition-recouvrement, et soustraction spectrale pour le rehaussement. Les résultats montrent que la méthode réduit le bruit mais introduit du noise music et dégrade le RSB. Des améliorations sont nécessaires : adaptation temporelle du seuil ou modèles perceptuels pour un rehaussement plus naturel.