Projet de traitement et filtrage numérique des signaux

Débruitage de parole par filtrage de Wiener avec modèle LPC

Introduction:

L'explosion du marché des télécommunications mobiles donne beaucoup de sens à l'intérêt porté à la problématique de débruitage de signaux de parole. Le bruit engendre en effet d'une part une dégradation de la qualité sonore du signal, mais surtout, d'autre part, une dégradation de l'intelligibilité du message transmis.

Le filtre de Wiener est le filtre qui fournit l'estimateur linéaire optimal d'un signal s(n) à partir d'une observation bruitée x(n) = s(n) + v(n), au sens de la MSE. Sa fonction de transfert découle alors du principe d'orthogonalité et des équations de Wiener-Hopf:

$$W(z) = \frac{\gamma_{sx}(z)}{\gamma_x(z)}$$

On se restreint ici au cas où *s(n)* et *v(n)* sont supposés décorrélés et où v(n) est blanc (DSP constante: σ_v^2).

Le codage par prédiction linéaire (ou LPC) est une méthode de codage et de représentation de la parole qui consiste à prédire de façon linéaire le signal à un instant n à partir de ses p échantillons précédents, de sorte à minimiser l'erreur de prédiction (due à la non-linéarité du signal). Sa fonction de transfert est la suivante:

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}$$

Cette représentation a l'avantage de pouvoir fournir des filtres numériques d'ordre raisonnable à partir d'une formule utilisant des DSP de signaux, comme c'est le cas pour le filtre de Wiener, c'est pourquoi nous nous y intéressons.

Objectifs du projet:

- Débruiter un signal de parole par filtrage de Wiener avec modélisation du signal par LPC;
- Recherche d'un traitement pseudo temps réel (Matlab ne permettant pas d'implémenter du vrai temps réel) afin de simuler au mieux un débruitage de parole en conditions réelles d'utilisation.

La plus grosse partie de l'étude est effectuée en premier lieu à l'échelle d'une trame d'un signal, sur laquelle on peut considérer que celui-ci est bien stationnaire. On étudie dans un second temps la généralisation de la méthode sur un signal complet, en mettant en oeuvre un filtrage sur des fenêtres d'analyse successives de courtes durées.

I/ Traitement d'une trame stationnaire (voyelle 'a' féminine):

L'idée naturelle est de commencer par implémenter le filtre de Wiener non-causal, dont on connaît plusieurs propriétés. On rappelle la fonction de transfert obtenue:

$$W(z) = \frac{G^2 / N\sigma_v^2}{A_1(z)A_1(z^{-1})} \quad \text{avec:} \quad A_1(z)A_1(z^{-1}) = G^2 / N\sigma_v^2 + A(z)A(z^{-1})$$

Ce filtre est donc un filtre à réponse impulsionnelle infinie, puisque tout pôles, et possède en plus, contrairement à ses homologues causaux, une phase nulle (sa fonction de transfert étant réelle positive) et donc un retard nul.

A/ Filtrage non-causal:

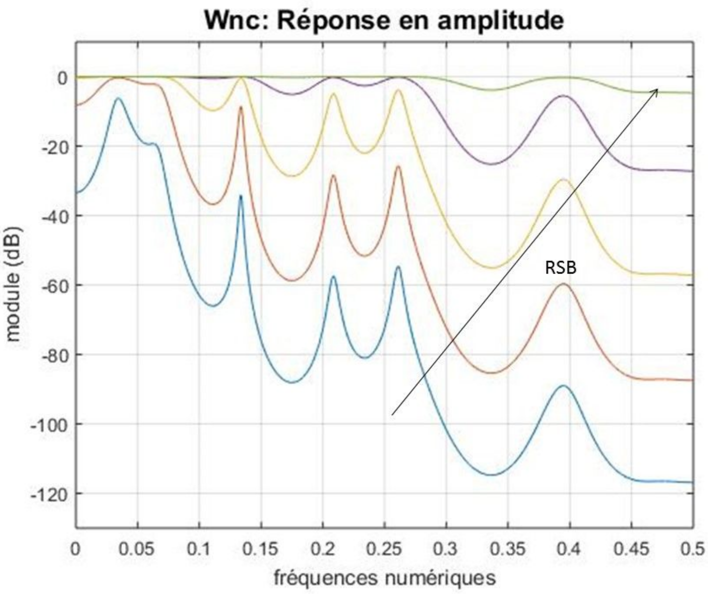
Matlab ne peut modéliser un filtre non-causal. Cependant ici, W(z) est un filtre symétrique, donc on a recours à un filtrage "forward-backward" à partir de la partie causale de ce filtre, et dont la sortie est rigoureusement identique à celle que donnerait le filtre non-causal.

W(z) se sépare en effet ici de manière évidente en un produit entre:

- une partie causale: $W_{c2}(z) = \frac{G / \sqrt{N\sigma_v^2}}{A_1(z)}$
- une partie anti-causale: $W_{a2}(z) = W_{c2}(z^{-1})$

On cherche ainsi à récupérer la partie causale de W(z) ici, à savoir $W_{c2}(z)$. Il s'agit donc de déterminer le polynôme $A_1(z)$ dont tous les zéros se trouvent à l'intérieur du cercle unité, et sont les symétriques aux zéros de $A_1(z^{-1})$ par rapport au cercle unité.

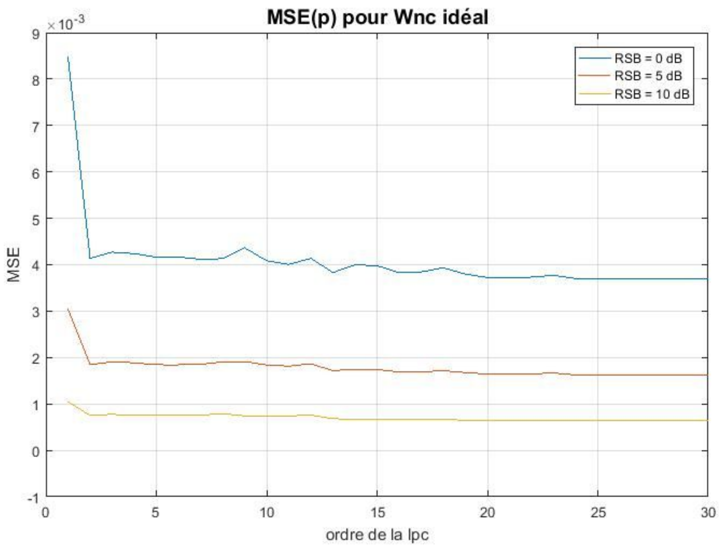
Influence du RSB (cas idéal, ordre LPC = 20, RSB de -15 dB à 45 dB):



Le filtre a donc bien la forme souhaitée:

- il agit comme un passe-bas lorsque le bruit est important, afin de l'éliminer, tout en conservant le signal de parole basse fréquence (typiquement compris entre 300 Hz et 5 kHz en général). On retrouve d'ailleurs bien des "pics", qui correspondent assez bien aux formants du signal du 'a' féminin étudié ici, aux fréquences: 750 Hz, 1390 Hz et 2950 Hz ;
- il s'aplatit petit à petit lorsque le RSB augmente pour finalement tendre vers un passe-tout lorsque le signal de parole prédomine largement sur le bruit. Il n'y a alors en effet plus aucun intérêt à le filtrer.

On peut ensuite montrer les courbes de la MSE en fonction de l'ordre de la LPC pour différents RSB:



L'erreur quadratique moyenne (ou MSE) obtenue semble donc suivre la même tendance quelque soit le RSB et atteint des valeurs de l'ordre de 10^{-3}. On peut toutefois remarquer une nette amélioration d'un facteur 4 entre le cas RSB = 0 dB et le cas RSB = 10 dB. De plus, et conformément à la théorie, la MSE diminue avec l'ordre de la LPC. On peut notamment constater que l'erreur diminue très fortement pour p variant de 1 à 3 et ne varie quasiment plus pour un ordre supérieur à 25. C'est pourquoi un ordre de LPC aux alentour de 15 semble être un bon compromis entre précision et temps de calcul.

Pour finir, on étudie le filtre avec la LPC estimée non plus à partir du signal non-bruité s(n) la source mais sur le signal bruité. Dans ce cas, les résultats dépendent du RSB initial du signal: dans un environnement trop bruité, la LPC sera calculée sur le bruit majoritairement et le résultat ne sera pas satisfaisant, malgré des filtrages successifs.

Cependant pour espérer utiliser le filtre en temps réel, il est nécessaire que celui-ci soit causal. C'est pourquoi l'étude des formes causales du filtre de Wiener fait l'objet de la partie suivante.

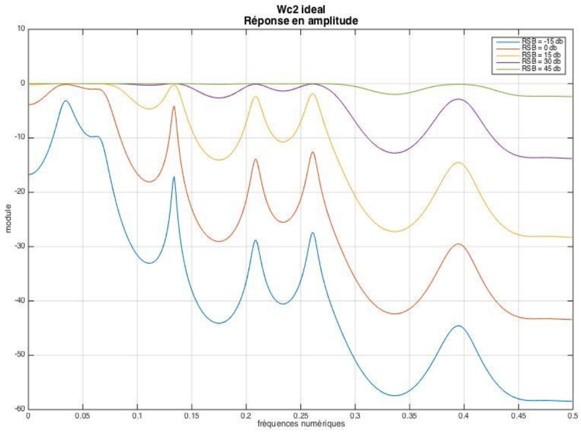
B/ Filtrage causal:

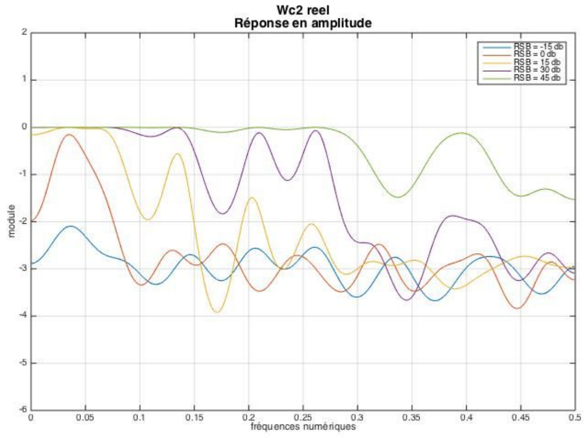
Les filtres étudiés ici étant causaux, le signal de sortie sera nécessairement retardé par rapport au signal d'entrée du filtre, et ce retard ne sera pas constant dans le cas des réponses impulsionnelles infinies.

- Forme "simple":

Cette forme correspond tout simplement à la partie causale de W(z) extraite précédemment, c'est à dire $W_{c2}(z)$. Néanmoins, ici, au lieu de réaliser un double filtrage comme dans la partie précédente, on réalise un unique filtrage, causal.

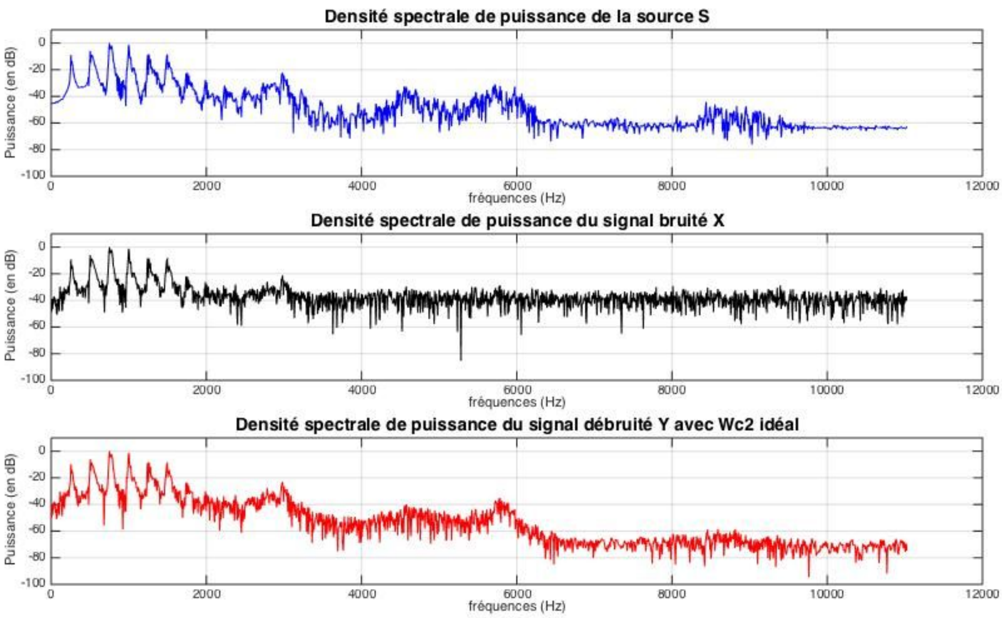
On observe tout d'abord la réponse fréquentielle en gain du filtre :





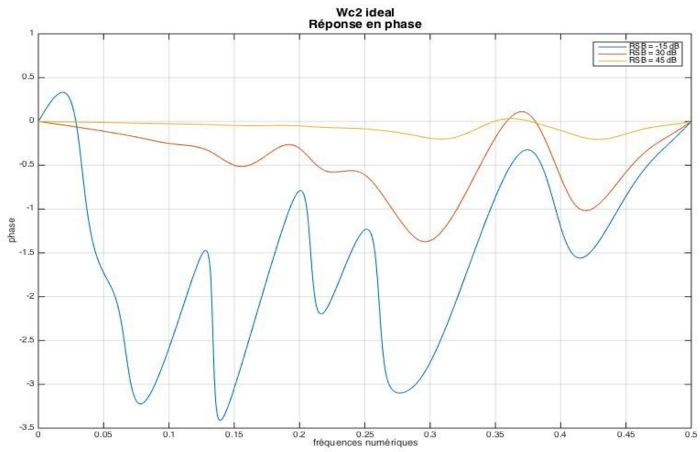
Pour la LPC calculée à partir de la source (cas idéal), on constate comme pour Wnc que la courbe de gain tend bien vers 0 dB lorsque le RSB augmente et se comporte comme un filtre passe-bas avec des pics au niveau des formants du 'a' féminin lorsque le RSB est faible. Pour le cas réel où la LPC est calculée à partir du signal bruité, le filtre tend bien vers 0 dB quand le RSB augmente mais ne se comporte plus vraiment comme un filtre passe-bas lorsque le RSB diminue, et les pics correspondant aux formants ne sont plus présents. Cela se traduit concrètement par une très bonne reconstruction à l'oreille du signal avec LPC idéale mais cette reconstruction semble bien moins bonne avec la LPC réelle.

Nous avons ensuite représenté et comparé les spectres des signaux S, X et Y :



On constate que les signaux S et X se différencient principalement par des composantes énergétiques hautes fréquences sur le signal bruité. Ces hautes fréquences dues au bruit sont enlevées par le filtre de Wiener qui agit comme un passe-bas et on vérifie bien ici que le spectre du signal débruité Y n'a plus de composantes énergétiques hautes fréquences (pour f >6 kHz).

On observe ensuite les réponses en phase de Wc2:



On constate premièrement que la phase est mal représentée par Matlab. En effet, la phase devrait être décroissante de 0 à Fe, et doit valoir pour f = Fe : $\phi = -n * 90°$, n étant le nombre de pôles du filtre. Cette mauvaise représentation est dûe à la définition première de la phase : $\phi = \arctan(\frac{Im}{Re})$ et arctan prenant ses valeurs dans [-90°,90°], on ne peut donc pas représenter naturellement la décroissance de la phase de 0° à -n*90°. On peut tout de même mentionner le fait que le filtre étant RII causal, il n'est pas à phase linéaire, ce qui implique des retards et des distorsions

Nous avons ensuite analysé la MSE du signal débruité Y par rapport à la source S en fonction de l'ordre de la LPC et nous avons constaté que l'erreur augmente avec l'ordre de la LPC. Cela s'explique justement par le fait que Wc2 n'est plus à phase constante et introduit un déphasage et donc un retard non nul sur le signal filtré. Il faudrait donc calculer la MSE entre la source et le signal filtré avancé du retard moyen, pour avoir une MSE qui traduit réellement la qualité du signal débruité que nous entendons. Malheureusement, nos tentatives n'ont pas données de courbes de MSE correctes (qui décroit avec p). Nous en avons déduit que l'approximation de la phase par une droite est trop grossière pour donner des résultats justes.

Pour conclure, si les résultats de Wc2 sont corrects, ce n'est plus un filtre de Wiener à proprement parler car il ne minimise plus la MSE.

- Forme réelle du Wiener causal:

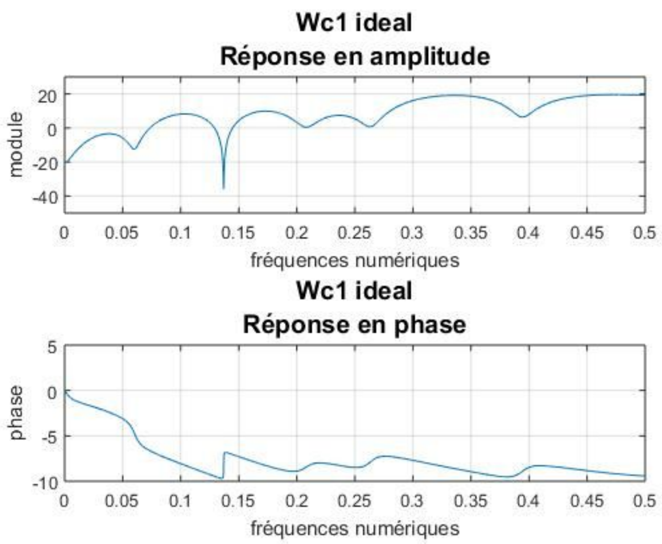
Le calcul théorique, basé sur l'annexe 2 et sur le cours de filtrage de SICOM 2A, nous amène à trouver la fonction de transfert suivante:

$$W_{c1}(z) = \frac{G^2}{N\sigma_v^2}\frac{B_1(z)}{A_1(z)} \qquad \text{où} \qquad B1(z) = \left\{\frac{1}{A(z).A1(z^{-1})}\right\}^{+} . A(z)$$

avec $\{ H \}^{+}$ qui correspond à la contribution causale de H dans sa décomposition en éléments simples.

La démarche suivie est la même que précédemment pour déterminer le polynôme A_1. La contribution causale de la fraction rationnelle est quant à elle déduite du résultat de la fonction *residuez* en ne conservant que les pôles situés à l'intérieur du cercle unité.

Malheureusement, en pratique, le filtre ne présente pas du tout la réponse fréquentielle attendue et semble même instable, malgré le fait que ses pôles soient tous contenus dans le cercle unité:

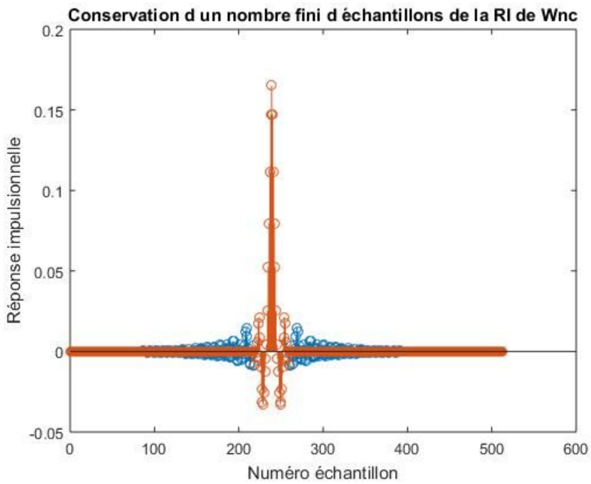


Ce filtre est censé minimiser la MSE tout en offrant un résultat sonore finalement moins bon que pour le filtrage causal "simple". En effet, pour respecter le principe de Wiener qui est de minimiser la MSE, le filtre doit normalement tenter ici de compenser le retard induit par sa causalité, qui est la principale source d'erreurs (en termes de MSE vis à vis du signal S) de W_{c2}. L'aspect amplitude du signal se retrouve alors en théorie "délaissé", contrairement au cas W_{c2} où il s'agissait du seul critère d'optimisation. Il ne fait rien de tout cela ici, de sorte que nous remettons en question la fonction de transfert théorique du filtre déterminée...

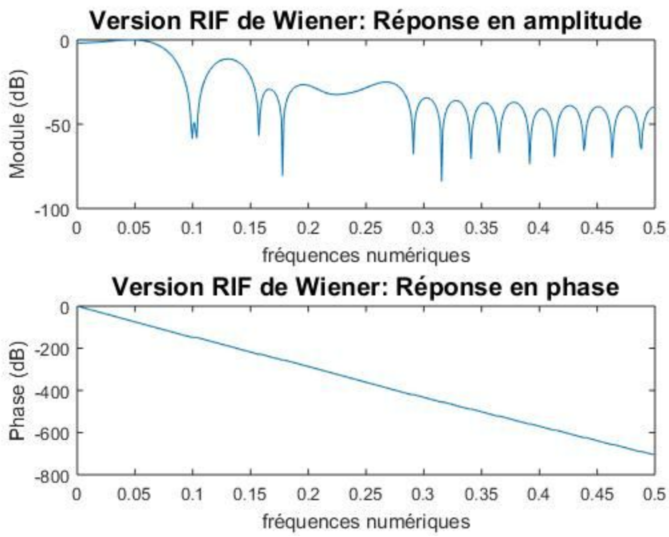
On voit quoiqu'il en soit que le déphasage non-linéaire engendré par les filtres de Wiener causaux pose des problèmes. Une idée intéressante est donc d'implémenter une version RIF du filtre de Wiener, causale, et à phase linéaire pour être en présence d'un retard constant et donc compensable.

- ## Version RIF causale du filtre de Wiener:

Pour réaliser un tel filtre, il suffit de conserver un nombre fini 2M+1 d'échantillons de la réponse impulsionnelle du filtre de Wiener non-causal de manière symétrique par rapport à l'échantillon n=0. C'est ce que représente la figure suivante pour une valeur de M = 20, dans le cas d'une estimation par LPC idéale et pour un RSB = 5 dB:



On déplace ensuite le résultat de M échantillons de sorte à rendre la réponse impulsionnelle causale. Le filtre obtenu est alors bien à phase linéaire, donc à retard constant:



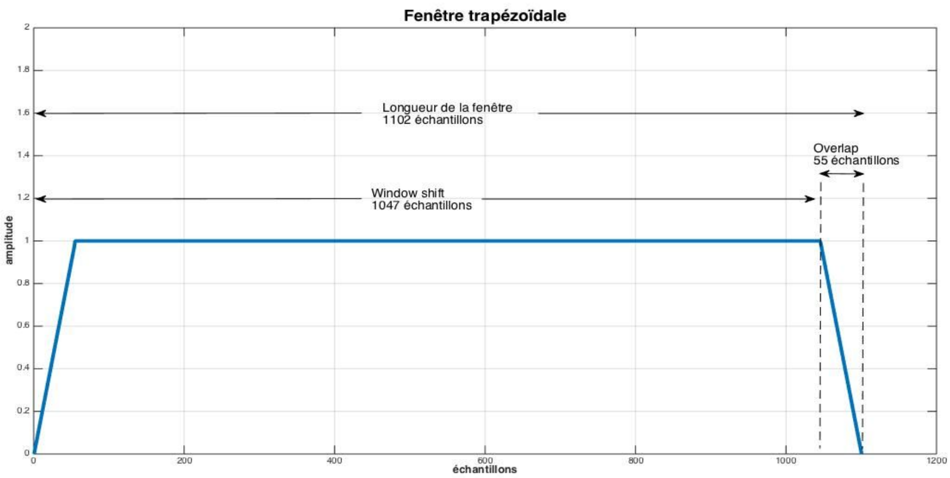
La courbe en amplitude présente des lobes aux mêmes fréquences que dans le cas du filtre de Wiener non-causal, mais ces lobes sont alors plus larges, et on voit surtout apparaître des lobes secondaires. Cela s'explique par le fenêtrage de type "porte" effectué sur la réponse impulsionnelle du filtre de Wiener non-causal, la transformée de Fourier d'une porte étant un sinus cardinal. Pour réduire l'importance de ces lobes secondaires, il pourrait être intéressant d'utiliser un fenêtrage plus "doux" de type Hanning par exemple.

Avec la fenêtre rectangle, le signal restitué en sortie est audible et très intelligible, même pour des RSB faibles comme 5 dB par exemple, dans le cas idéal.

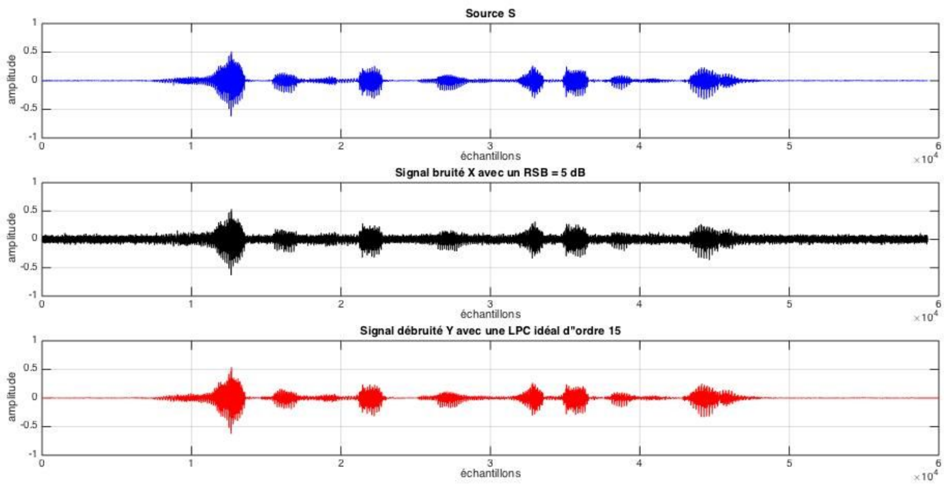
II/ Traitement sur un signal de parole entier:

Dans cette partie, nous allons voir comment utiliser ces différents filtres sur des signaux de parole entiers. Comme ces signaux sont trop longs pour être considérés stationnaires, on ne peut pas appliquer directement ces filtres sur tout le signal. Il s'agit donc d'appliquer la méthode d'overlap-add : on découpe le signal par tranche, on filtre chaque partie tronquée puis on les pondère par une fenêtre trapézoïdale. Enfin, on somme les parties filtrées et pondérées en les superposant en partie pour sommer les résidus de chaque filtrage.

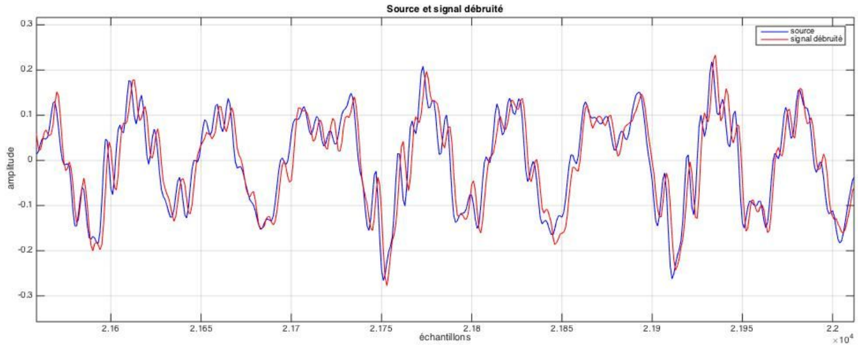
Dans un premier temps, nous avons filtré le signal '101' qui correspond à la phrase : "Guy a péri bêtement du diabète en Italie" avec une fenêtre trapézoïdale de 50 msec (1102 échantillons) et d'un overlap sur 2.5 msec (55 échantillons).



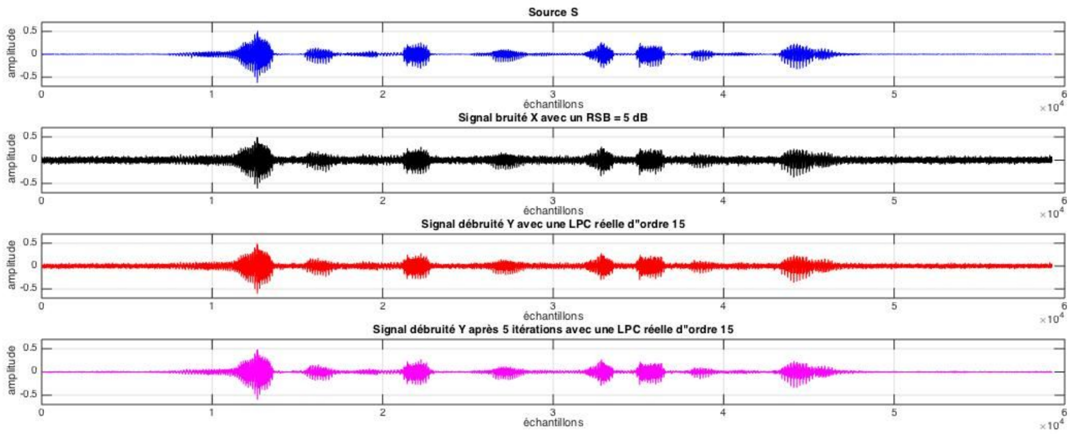
On obtient les résultats suivants pour une LPC idéale :



On constate à l'oreille que le bruit de fond a totalement disparu, en revanche le signal de parole semble avoir subi des distorsions, et bien qu'il soit toujours compréhensible, il n'est pas très agréable à écouter. Cela ne se voit pas sur les courbes temporelles qui montrent pourtant un signal débruité qui suit bien le signal original malgré le léger retard dû au déphasage de Wc2 (comme vu précédemment, l'oreille ne perçoit pas ce déphasage).



Pour le cas réel où la LPC est calculée à partir du signal bruité, le résultat est beaucoup moins bon à l'oreille, le signal débruité Y ressemble beaucoup au signal X avec un niveau de bruit légèrement moins fort. C'est pourquoi on utilise la méthode de Lim et Oppenheim qui consiste à refiltrer Y plusieurs fois. Après plusieurs itérations, on s'aperçoit en effet que le signal débruité présente de moins en moins de bruit avec en revanche une distorsion du signal de parole comme pour le cas idéal:



Il s'agit ici de trouver un compromis satisfaisant : si le nombre d'itérations est trop faible, le signal débruité Y est très proche du signal bruité X. Si le nombre d'itérations est trop grand, le bruit a complètement disparu mais le signal de parole n'est plus compréhensible à cause des distorsions importantes qu'ont généré les nombreux filtrages successifs.

Pour finir, nous avons filtré le même signal avec cette fois la version RIF du filtre de Wiener. Les résultats sont à l'oreille aussi bons qu'avec Wc2. Cette version est donc une très bonne alternative puisque le RIF a l'avantage de représenter moins de calculs et peut ainsi être facilement implémenté sur DSP.

Conclusion :

Les résultats obtenus sont donc concluants puisqu'ils montrent que le filtre de Wiener sous sa forme causale permet de retirer significativement la présence de bruit tout en gardant la structure essentielle du spectre de parole et donc conserver l'information. Nous avons également montré que l'utilisation de Wc2 sur tout un signal avec un filtrage par overlap-add permettait de filtrer de manière correcte une grosse partie du bruit, grâce à la méthode de réitération de Lim et Oppenheim. Un point qui reste à améliorer est la formule de Wc1 qui manifestement pose problème. Ce filtre causal plus complexe que Wc2 pourrait potentiellement présenter de meilleurs résultats que Wc2 sur un signal de parole réel après fenêtrage et serait donc une solution intéressante pour un problème concret de filtrage en temps réel.