Travaux Pratiques Traitement Numérique du Signal

FILIERE ELECTRONIQUE, 2^E ANNEE

2026-2027

Rémi Giraud
remi.giraud@enseirb-matmeca.fr

Éloi Navet
eloi.navet@u-bordeaux.fr

TABLE DES MATIÈRES

1 ANNEXES **3**
1.1 Séparation de sources 3
1.2 Notion de bruit blanc gaussien 4
1.3 Débruitage d'un signal de parole 5
1.3.1 Caractérisation des voyelles et notions de formants 6
1.3.2 La fréquence fondamentale 6
1.3.3 Bilan 7
1.3.4 Sur le rehaussement monovoie par atténuation spectrale à court terme . 7
1.4 Formulaire 9
1.5 Mémo Matlab 10

1 ANNEXES

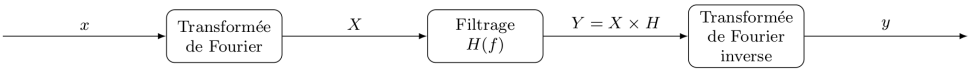
1.1 Séparation de sources

Un des grands domaines du traitement du signal est la séparation de sources. L'idée étant, à partir d'un mélange de différents signaux provenant de plusieurs sources (par exemple plusieurs personnes qui parlent en même temps), de parvenir à reconstruire le message associé à chacune des sources (par exemple retrouver ce qu'a dit chaque personne). C'est un problème courant en acoustique donc, mais aussi en télécommunications, en bio ingénierie, etc.

Dans notre cas, nous allons considérer la somme de deux signaux sonores (mélodies), et tenter de reconstruire chacun d'entre eux grâce à un filtrage adapté, directement dans l'espace des fréquences. Pour cela, nous allons utiliser au choix un filtre passe-bas idéal (si l'on veut supprimer les notes aiguës), passe-haut idéal (si l'on veut supprimer les notes graves). On peut aussi considérer un passe-bande (et ainsi choisir de supprimer toutes les notes aiguës et graves et ne garder que les médiums). Dans cette partie, le signal est composé de 2 mélodies sur des gammes de fréquences différentes. En pratique, la séparation de sources sur des mêmes espaces fréquentiels peut s'avérer beaucoup plus complexe. Les fonctions de transfert de ces filtres sont définies de la façon suivante et dépendent d'une ou plusieurs fréquences de coupure (selon le filtrage que l'on souhaite effectuer) :

Passe-bas	Passe-haut	Passe-bande
$H_{LP}(f) = \begin{cases} 1 & \text{si } \lvert f\rvert < f_c \\ 0 & \text{sinon} \end{cases}$	$H_{HP}(f) = \begin{cases} 0 & \text{si } \lvert f\rvert < f_c \\ 1 & \text{sinon} \end{cases}$	$H_{BP}(f) = \begin{cases} 1 & \text{si } f_{c1} < \lvert f\rvert < f_{c2} \\ 0 & \text{sinon} \end{cases}$

La suite des opérations à réaliser est présentée ci-dessous :



Il s'agit ici d'un filtrage linéaire vu dans le domaine fréquentiel, correspondant donc à une multiplication de la transformée du signal avec la fonction de transfert du filtre.

1. Avec la fonction *create_melody*, créer deux mélodies x_1 et x_2 d'une durée de 10 secondes, échantillonnés à $f_e = 8000$Hz et contenant des notes (au moins 4), pour x_1 comprises entre 50 et 71 et pour x_2 comprises entre 72 et 90.

   Écouter les signaux et visualiser leurs transformées de Fourier.

2. Créer le signal $x = x_1 + x_2$.

   Écouter le signal et visualiser sa transformée de Fourier.

Quelle fréquence de coupure f_c peut-on utiliser pour séparer les deux signaux x_1 et x_2 ?

3. Créer un filtre adapté (vecteur de 1 et de 0, de la taille de la transformée de Fourier), multiplier ce vecteur à la transformée du signal et utiliser la transformée inverse (*ifft*) pour reconstruire à partir de x les signaux y_1 et y_2 estimant respectivement x_1 et x_2.

   Quelle hypothèse faut-il sur M pour pouvoir effectuer cette transformée inverse ?

4. Écouter les signaux obtenus et les comparer aux sources réelles.

   Calculer la distance euclidienne entre x_1 et y_1 et celle entre x_2 et y_2.

5. Charger le signal *melody_multiple.wav*. En utilisant l'information du spectrogramme, essayer d'appliquer la stratégie précédente pour séparer la mélodie du fond.

   On pourra calculer la distance euclidienne entre le signal bruité et l'original contenu dans *melody_single.wav*.

1.2 Notion de bruit blanc gaussien

Une séquence Gaussienne de Bruit Blanc Gaussien Centré a trois propriétés caractéristiques :

Blanc
La notion de blancheur implique que tous ses échantillons sont décorrélés, *i.e.*,

$$R_{BB}(t_i, t_j) = \sigma^2 \delta_{t_i,t_j} \tag{1}$$

où δ_{t_i,t_j}, le symbole de Kronecker, est tel que δ_{t_i,t_j} si $t_i \neq t_j$ et $\delta_{t_i,t_j} = 1$ si $t_i = t_j$. De manière général, un processus blanc est stationnaire. On a donc $R_{BB}(l) = \sigma^2 \delta_l$.

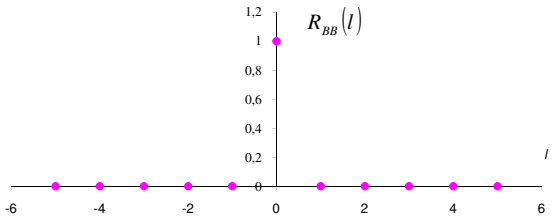


Figure 1 : Fonction d'autocorrélation du bruit blanc

Centré
Sa moyenne est nulle.

Gaussien
Un processus aléatoire est un processus gaussien si pour tout ensemble d'instants $\{t_i\}$, le vecteur ligne aléatoire correspondant $X = [X_1, \ldots, X_n]$, a une densité de probabilité conjointe multidimensionnelle gaussienne $X_i = X(t_i)$. Chaque X_i possède une densité de probabilité marginale :

$$p_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}\sigma_{X_i}} \exp\left(\frac{-[x_i - m_{x_i}]^2}{2\sigma_{X_i}^2}\right). \tag{2}$$

Si le processus est réel et $N = 1$ (cas mono-dimensionnel), sa densité de probabilité est définie par :

$$p(x;t) = \frac{1}{\sqrt{2\pi}\sigma_x^2(t)} \exp\left(\frac{-[x - m_x(t)]^2}{2\sigma_x^2(t)}\right). \tag{3}$$

Si le processus est complexe et $N = 1$, sa densité de probabilité est définie par :

$$p(x;t) = \frac{1}{\pi\sigma_x^2(t)} \exp\left(\frac{-[x - m_x(t)]^2}{\sigma_x^2(t)}\right). \tag{4}$$

Si le processus est réel et de dimension N, sa densité de probabilité est définie par :

$$p(x;t) = \frac{1}{(\sqrt{2\pi})^N |C_{XX}|^{1/2}} \exp\left(\frac{-1}{2}[x - m_x(t)]C_{XX}^{-1}[x - m_x(t)]^T\right). \tag{5}$$

Si le processus est complexe et de dimension N, sa densité de probabilité est définie par :

$$p(x;t) = \frac{1}{\pi^N |C_{XX}|} \exp\left([x - m_x(t)]C_{XX}^{-1}[x - m_x(t)]^T\right). \tag{6}$$

C_{XX} est la matrice d'autocovariance.

1.3 Débruitage d'un signal de parole

Le rehaussement du signal de parole consiste en l'amélioration d'un ou de plusieurs aspects perceptifs du signal vocal quand ce dernier est perturbé par un bruit dû à l'environnement et/ou à la transmission. Un traitement de débruitage est ainsi souhaitable dans une voiture, lors de conversations main libre, dégradées par le bruit du moteur et les conditions du trajet (pluie, fenêtre ouverte, etc.). La situation est comparable dans le domaine de l'aéronautique, lors de communications entre le pilote et la tour de contrôle. C'est aussi le cas dans les situations suivantes :

- les postes de téléphone situés dans des environnements où les bruits des conversations et des équipements informatiques perturbent les communications ;
- les vidéophones et systèmes de téléconférence ;
- la restauration d'enregistrements anciens ;
- les systèmes de reconnaissance utilisés dans des environnements bruités.

Nous supposons ici qu'un seul microphone est utilisé pour la prise de son et nous cherchons à réduire le bruit sans introduire de distorsion trop importante sur le signal utile.

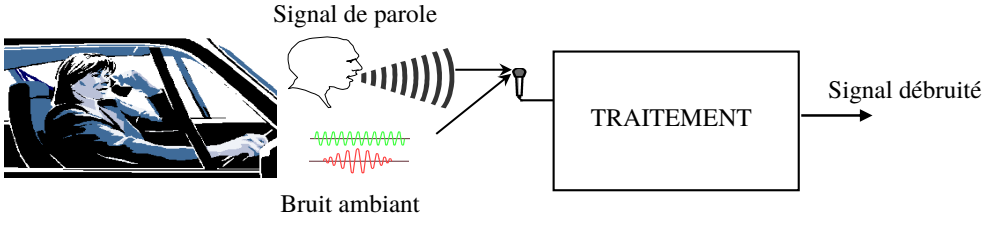


Figure 2 : Débruitage du signal de parole

Nous nous placerons dans le cas simple où le bruit additif est un bruit blanc.

Ce document s'organise de la manière suivante : dans un premier temps, quelques notions sur la parole sont données telles que les notions de voisement/non voisement, fréquence fondamentale, etc. Puis, une présentation théorique de l'approche de rehaussement est réalisée.

1.3.1 Caractérisation des voyelles et notions de formants

Par essence, le signal de parole est *non stationnaire* ; cependant, on peut faire une hypothèse de quasi-stationnarité, notamment pour des voyelles, sur une durée de 40 ms. En fenêtrant le signal de parole en trames de 10 à 40 ms, on peut faire apparaître un caractère pseudopériodique ou/et aléatoire. Plusieurs cas peuvent être envisagés :

les trames **voisées** qui présentent un caractère périodique ou pseudo périodique ;

les trames non voisées ou aléatoires ;

les "bruits voisés" qui sont une combinaison simultanée des deux cas précédents ;

les trames **mixtes** représentant les zones de transition entre des sons voisé et non voisé.

Le spectre d'un signal voisé est riche en fréquences harmoniques de la fréquence fondamentale. Si l'on regarde la figure 2, on constate en outre que l'enveloppe spectrale du signal analysé décroît doucement vers les hautes fréquences. Ses résonances sont appelées *formants* et notées $F1$, $F2$, etc. dans l'ordre des fréquences croissantes. Les deux premiers caractérisent en général la voyelle prononcée alors que les deux suivants sont plutôt liés au locuteur. Pour un locuteur masculin, tous les formants sont inférieurs à 5 kHz.

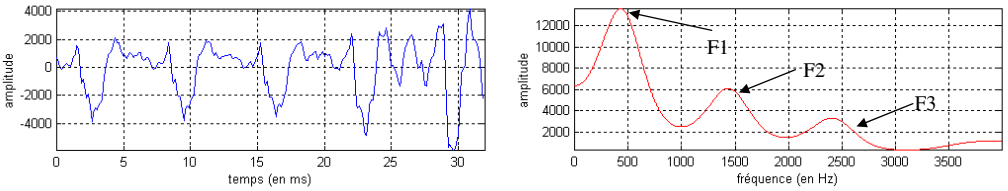


Figure 3 : Mise en évidence des trois premiers formants pour le segment voisé analysé

On peut aussi observer les formants sur un spectrogramme. Le spectrogramme est une représentation graphique fréquence×temps, qui traduit l'évolution de la richesse spectrale du signal au cours du temps ; la présence de formants correspond alors à ces gammes de fréquence dont l'énergie est particulièrement élevée et qui apparaissent sous la forme de bandes sensiblement parallèles à l'axe des abscisses. Quand le spectrogramme est en couleur, on adopte généralement une palette allant du rouge pour représenter une forte énergie au bleu pour représenter une faible énergie. C'est notamment le cas à la Figure 4.

1.3.2 La fréquence fondamentale

Lorsque les cordes vocales oscillent de façon quasi-périodique, le signal de parole émis est lui aussi quasi-périodique ; il est dit voisé, c'est notamment le cas lorsque l'on prononce des voyelles telles que le /a/. La période du signal est alors appelée période fondamentale. On nomme l'inverse de cette période, la fréquence fondamentale ou pitch. La valeur moyenne du pitch est variable selon les locuteurs. Basse et de l'ordre de 80 à 200 Hz pour les hommes, elle peut atteindre 600 Hz pour les enfants, la voix féminine ayant une fréquence fondamentale comprise de manière générale entre 150 et 450 Hz. Un exemple de spectrogramme d'un signal de parole[1] est donné en Figure 4.

[1]Nous remercions Matra et l'ENST pour la base de données de signaux audio qu'elles ont bien voulu nous fournir.

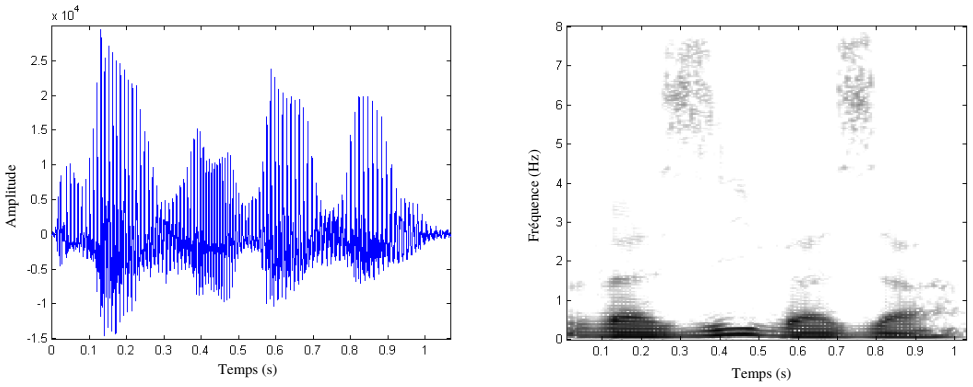


Figure 4 : Spectrogramme d'un signal "waziwaza"[1] mettant en évidence la présence de $[z]$

1.3.3 Bilan

On retiendra de cette brève présentation du signal de parole deux de ses caractéristiques majeures :

son caractère non-stationnaire (nécessitant un traitement par frames),

et son éventuelle nature quasi-périodique, à laquelle se rapportent les notions de voisement et de non voisement.

La parole est un moyen de communication au même titre que l'écriture. On attend donc d'un message vocal qu'il soit intelligible, c'est à dire facile à comprendre.

Mais le signal de parole peut être dégradé par un bruit additif, notamment lors de communications longue-distance sur des lignes bruitées, ou lorsque l'on utilise un téléphone portable dans une voiture. Des traitements de débruitage du signal peuvent alors être mis en œuvre. Dans le chapitre suivant, nous nous proposons de faire l'état de l'art des méthodes de rehaussement qui utilisent un unique microphone.

1.3.4 Sur le rehaussement monovoie par atténuation spectrale à court terme

Les solutions élaborées pour le débruitage sont multiples. Certaines emploient plusieurs microphones ; on peut effectuer un rehaussement bivoie, en plaçant un premier microphone directionnel à proximité du locuteur alors qu'un second microphone, omnidirectionnel, est positionné le plus près possible de la source de bruit. Ici, nous focalisons notre attention sur une approche de rehaussement de type monovoie. Une des difficultés dans ce contexte est d'obtenir une bonne approximation des données statistiques du bruit perturbateur.

Les techniques de rehaussement par atténuation spectrale à court terme consistent à appliquer une atténuation spectrale, dépendant de l'estimation du bruit additif, à chaque composante fréquentielle. Plus précisément, ces méthodes nécessitent les opérations suivantes :

1. On effectue un traitement tous les N échantillons. La durée des tranches d'analyse est comprise entre 20 et 40 ms. Les blocs que l'on traite successivement peuvent être en outre disjoints ou se recouvrir partiellement ; on adopte d'ailleurs couramment un recouvrement de 50%. Dans le contexte de la téléphonie, où le signal est échantillonné à 8 kHz, on peut travailler sur des blocs de 256 échantillons dont le recouvrement s'élève à 128 échantillons.

2. Puis, on applique une règle de suppression de bruit au spectre du signal bruité fenêtré $Y_w(f)$, que l'on nomme aussi spectre du signal à court terme, pour en atténuer chaque canal fréquentiel et obtenir ainsi $\hat{S}(f)$, l'estimation du spectre à court terme du signal de parole original :

$$\hat{S}_w(f) = H_w(f, B(f))Y_w(f) \tag{7}$$

Cette règle varie selon les méthodes utilisées et dépend de manière générale de la variance du bruit additif dont on fait une estimation préalable ; on peut tirer avantage des instants de "silence", où seul le bruit additif est présent, pour évaluer et/ou actualiser la variance du bruit. On en fait alors une approximation long terme, en supposant le bruit ergodique[2].

3. Une transformation de Fourier inverse permet enfin d'avoir une estimation du signal fenêtré. Puis on procède à la reconstruction du signal, en appliquant une méthode dite d'addition-recouvrement (addition de l'ensemble des signaux fenêtrés débruités).

   Le schéma de principe de rehaussement par atténuation spectrale est donné à la Figure 6.

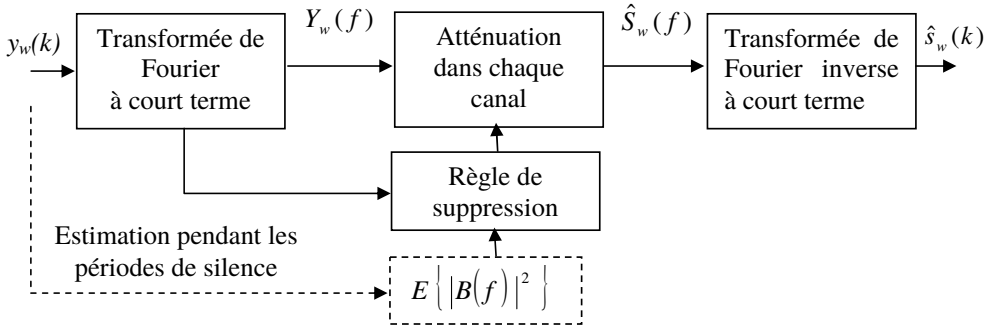


Figure 5 : Schéma de principe de rehaussement par atténuation spectrale

Parmi les méthodes qui obéissent à ce type de traitement, la soustraction spectrale est l'une des techniques les plus employées par les traiteurs de signal de parole.

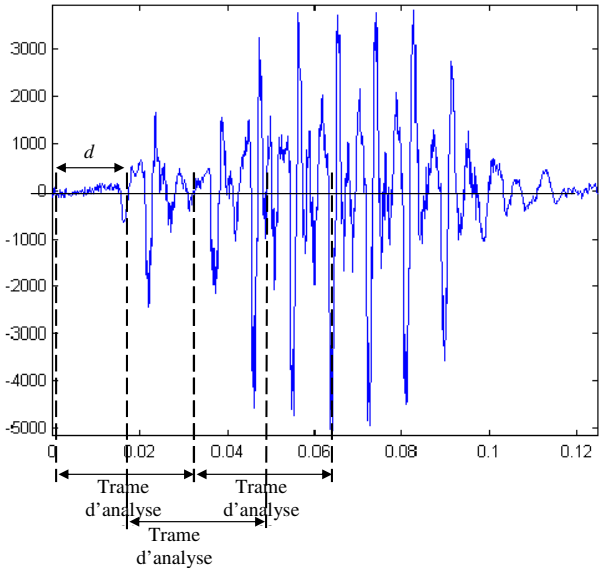


Figure 6 : Traitement par trame avec recouvrement d'un signal

[2]"èrgodique au sens de la moyenne" revient à dire que l'on peut approximer la moyenne d'ensemble du processus, c'est à dire l'espérance mathématique par la moyenne temporelle.

1.4 Formulaire

Domaine continu	Domaine discret
$X(p) = \int_{-\infty}^{+\infty} x(t)\exp(-pt)dt$	$X(z) = \sum_{n=-\infty}^{+\infty} x(n)z^{-n}$
Équation différentielle à coefficients constants $\sum_{k=0}^{K-1} C_k \frac{d^k y(t)}{dt^k} = \sum_{l=0}^{L-1} D_l \frac{d^l x(t)}{dt^l}$	Équation récurrentes $\sum_{k=0}^{K-1} a(k)y(n-k) = \sum_{l=0}^{L-1} b(l)x(n-l)$
fréquence : $p = jw = j2\pi f$	fréquence : $z = \exp(jwT_e) = \exp(j2\pi\frac{f}{f_e})$
Fonction de Transfert $H(p) = \frac{\sum_{l=0}^{L-1} D_l p^l}{\sum_{k=0}^{K-1} C_k p^k}$	Fonction de Transfert $H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{l=0}^{L-1} b(l)z^{-l}}{\sum_{k=0}^{K-1} a(k)z^{-k}}$
Stabilité (système causaux) : pôles à partie réelle négatives	Stabilité (système causaux) : pôles à l'intérieure du cercle unité

On rappelle en outre que :

$$\sum_{n=0}^{N-1} a^n = \frac{1-a^N}{1-a} \quad \text{d'où} \quad \sum_{n=0}^{+\infty} a^n = \frac{1}{1-a} \quad \text{pour} \quad |a| < 1$$

1.5 Mémo Matlab

```
%Chargement d'un signal 'àudio''
[s, fe] = audioread('signal.wav');
N = length(s)

%Chargement d'une archive matlab
load('data.mat'); %Données chargées directement dans le workspace
X = load('data.mat'); %Données dans X
s = X.s; %Accès aux champs de X

%Valeur maximale
max_s = max(s);
%Accès valeurs
s_10_20 = s(10:20); valeurs 10 à 20

%Mise à zéro
s = zeros(N,1); %ones() existe aussi
s = s*0;
s(5:20) = 0; %mise à 0 des coeff 5 à 20

%Vecteurs/matrices
vect = (1:2:11); %(début:(pas):fin) [1 3 5 7 9 11] (1x6)
vect = linspace(1,11,6); %(début,fin,N) [1 3 5 7 9 11] (1x6)
mat = [1 1; 2 2; 3 3]; %matrice (3x2)
vect_t = vect'; %transposée (6x1)

%Attention aux * de vecteurs et matrices !!!
vect_x = vect*vect; %ERREUR DE DIMENSION
vect_1x6 = vect.*vect; %terme à terme
vect_6x6 = vect'*vect; %produit matriciel
vect_1 = vect*vect'; %produit matriciel

%Sous-échantillonnage
s_ech = s(1:2:end); %une valeur sur deux

%Somme des termes au carrés
sum_s = sum(s.^2); %équivalent à sum(s.*s);

%Boucles
vect = [2,4,5,18,9];
N = length(ind);
for ind = 1:N
     v = vect(ind);
     ...
end

%Création d'un cosinus
N = 1024; t = 0:N-1; %de longueur N
fe = 256; f0 = 64; %freq ech fe, porteuse f0
s = cos(2*pi*f0/fe*t);
```

```
%Affichage
figure, %Création d'une nouvelle figure
subplot(121) %Affichages multiples 1x2
plot(1:length(s), s) %première figure
title('Signal s');
subplot(122) %deuxième figure
plot(s); %indices 1 à length(s) par défaut
hold on; superposition
plot(1:length(s2), s2, 'Color', [0 0 1]); title('Comparaison de s et s2'); xlabel('Signal');
ylabel('Intensité');
```