

# Débruitage audio

Matthieu Kowalski

Univ Paris-Sud – L2S (GPI)

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Origine des "bruits"

- Conditions de prises de sons (Concert en plein air)
- Bruits environnants (Applaudissements, personne qui tousse)
- Bruits électroniques (le "50 Hz")
- Saturation
- Dégradation de la qualité par codage (MP3)
- Dégradation des enregistrements
- etc.

# Bruits considérés

- Bruits de fonds
- Craquements / clicks
- (Données manquantes)

Comment compenser ces dégradations ?

# But de la restauration

Point de vue d'un ingénieur du son

*"Tenter de retrouver le signal original n'a aucun sens. Il faut produire un nouveau signal de qualité acceptable"*

Comment évaluer la qualité de restauration ?

# Évaluation

Il faut distinguer deux types d'évaluations :

- ① Le point de vue de l'ingé son : il faut produire un signal de qualité acceptable à partir d'un enregistrement. **Mesure subjective** par définition.
- ② Le point de vue du traiteur de signal : concevoir des algorithmes qui fonctionnent. Phase de test des algorithmes avec des **mesures objectives**.

# Position du problème

On se limitera aux bruits additifs :

$$\mathbf{y} = \mathbf{x} + \mathbf{n}$$

avec

- $\mathbf{y} \in \mathbb{R}^T$  le signal mesuré (de longueur  $T$  échantillons)
- $\mathbf{x} \in \mathbb{R}^T$  le signal original (à estimer)
- $\mathbf{n} \in \mathbb{R}^T$  un bruit additif

But

Estimer  $\mathbf{x}$  à partir des seules mesures  $\mathbf{y}$

# Plan

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Outline

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Définition

## Définition

Un bruit de fond est un processus stationnaire.

## Exemples

- Les "souffles"
- Les applaudissements
- bruits de roulements
- etc.

D'un point de vue modélisation statistique :

- Bruit blanc (processus gaussien)
- Bruit coloré (processus gaussiens)
- Tout processus stationnaire (définition)

# Rappels

## Processus stationnaire

Un processus est dit stationnaire si ses statistiques (moyenne, variance, etc) sont invariantes par translation.

## Densité spectrale de puissance et théorème de Wiener-Khintchine

Soit  $X$  un processus stationnaire. La densité spectrale de puissance  $S_X$  est égale à la transformée de Fourier de la fonction d'autocorrélation  $R_X$ .

$$S_X(f) = \int_{-\infty}^{+\infty} R_X(t) e^{-i2\pi ft} dt$$

On rappelle que :

$$S_X(f) = \mathbb{E} [|\hat{X}(f)|^2] \quad R_X(t) = \mathbb{E} [x(\tau)x(\tau + t)]$$

# Bruits blancs

## Bruit blanc fort

Soit  $X = (X_k)_{k \in \mathbb{N}}$  un processus aléatoire discret.  $X$  st un bruit blanc fort si les  $X_k$  sont centrés et i.i.d.

## Bruit blanc faible

Soit  $X = (X_k)_{k \in \mathbb{N}}$  un processus aléatoire discret.  $X$  st un bruit blanc faible si les  $X_k$  sont centrés, décorrélés et de variance finie.

## Remarque

Pour les processus gaussien, les notions de bruits blancs forts et faibles sont confondues

# Bruits blancs gaussiens

## Théorème : bruit blanc gaussien

Soit  $X = (X_k)_{k \in \mathbb{N}}$  un processus aléatoire discret.  $X$  est un bruit blanc gaussien ssi

- Les  $X_k$  sont indépendants
- $X_k \sim \mathcal{N}(0, \sigma_X^2)$

## Propriétés

- La fonction d'autocorrélation est un dirac :

$$R_X(t) = \delta(t)$$

- La densité spectrale de puissance de  $X$  est constante :

$$S_x(f) = \sigma_X^2$$

# Bruits gaussien colorés

## Définition

$X$  est un bruit gaussien coloré si  $X$  est un processus gaussien centré dont la densité spectrale de puissance n'est pas constante

## Exemples

- Bruit rose :  $S_X(f) = \frac{\alpha}{f}$ ,  $\alpha > 0$
- Bruit rouge/brun/brownien :  $S_X(f) = \frac{\alpha}{f^2}$
- Bruit bleu :  $S_X(f) = \alpha f$
- Bruit violet :  $S_X(f) = \alpha f^2$

# Bruit gaussien non blanc

- Les bruits blancs et colorés sont des modèles théoriques.
- En pratique, on estime la densité spectrale du bruit  $S_X$  qui peut être quelconque.

## Propriété

Soit  $X$  un processus gaussien et  $h$  un filtre. Soit

$$Y = h \star X$$

Alors

$$S_Y(f) = |h(f)|^2 S_X(f)$$

# Bruit gaussien non blanc

## En particulier

Soit  $X$  un processus gaussien de densité  $S_X(f)$ . Alors  $X$  peut être obtenu par filtrage d'un bruit blanc gaussien normé :

$$X = h \star N, \quad N \sim \mathcal{N}(0, 1)$$

avec  $S_X(f) = |h(f)|^2$

On a une manière simple générer n'importe quel type de bruit gaussien

# Modèle des bruits de fond

On considèrera le problème :

$$\mathbf{y} = \mathbf{x} + \mathbf{n}$$

où

- $\mathbf{n}$  est un bruit gaussien de densité spectrale inconnue (De-hissing)
- $\mathbf{n}$  est un bruit supposé seulement stationnaire (ex : applaudissements)

# Plan

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Outline

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Définition et modèle

- Les "clicks" : bruit impulsionnel de très courte durée
- Les "craquements" : bruit impulsionnel de durée plus longue

## Modèle

$$y_t = x_t + i_t n_t$$

- $n_t$  est un processus aléatoire
- $i_t = 0/1$  un processus aléatoire binaire :  $i_t = 1$  lorsque le bruit est présent.

# Modèles aléatoires

## Modèles simples

- Processus Bernoulli-Gaussien :  $\mathbf{i} \sim \mathcal{B}(p)$  et  $\mathbf{n} \sim \mathcal{N}(0, \Sigma)$
- Processus Poisson-Gaussien :  $\mathbf{i} \sim \mathcal{P}(\lambda)$  et  $\mathbf{n} \sim \mathcal{N}(0, \Sigma)$

Peut être satisfaisant pour les clics, mais pas pour les craquements

## Modèles plus évolués

- Modéliser la durée du craquement : dépendances fortes entre les échantillons successifs de  $i$
- Les amplitudes peuvent varier très fortement (et être très grande par rapport au signal original)
- Modèle de bruit transitoire par template

# Contexte

- 1 Mesures objectives (SNR)
- 2 Protocole d'évaluation subjective (MOS, MUSHRA)
- 3 Métrique perceptives PEAQ

# Mesure objective : SNR

Modèle de bruit additif général :

$$\mathbf{y} = \mathbf{x} + \mathbf{n}$$

Définition : SNR d'entrée

Le rapport signal à bruit en entrée défini par :

$$\text{SNR}(\mathbf{y}, \mathbf{x}) = 10 \log \left( \frac{\|\mathbf{y}\|^2}{\|\mathbf{y} - \mathbf{x}\|^2} \right)$$

Définition : SNR en sortie

Soit  $\hat{\mathbf{x}}$  une estimation de  $\mathbf{x}$ . Le SNR en sortie est donnée par

$$\text{SNR}(\hat{\mathbf{x}}, \mathbf{x}) = 10 \log \left( \frac{\|\mathbf{x}\|^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|^2} \right)$$

- Nécessite de connaître le "vraie" signal  $\mathbf{x}$  !
- Utile pour évaluer un algorithme sur des signaux test

# Métrique perceptive : PEAQ

## PEAQ

- Perceptual Evaluation of Audio Quality
- Mesure objective prenant en compte les propriétés connues de l'oreille humaine (masquage temps-fréquence)

# Protocole d'évaluation subjective

Deux protocoles principaux :

- 1 MOS : Mean Opinion Score
- 2 MUSHRA : MULTiple Stimuli with Hidden Reference and Anchor

## MOS

$N$  évaluateurs donne une note  $R_n$  entre 1 (Mauvais) et 5 (Excellent)

$$\text{MOS} = \frac{\sum_{n=1}^N R_n}{N}$$

## MUSHRA

- Élaboré pour évaluer les codeurs avec perte. Toutes les versions compressées sont présentées en même temps, sur les mêmes signaux à tous les utilisateurs.
- Nécessite moins de participants que MOS.

# Problèmes

$$\mathbf{y} = \mathbf{x} + \mathbf{n}$$

avec  $\mathbf{n}$  bruit gaussien, de densité spectrale  $S_n(f)$  inconnue.

**Hypothèse** : on peut observer une partie du bruit "seul"  $\mathbf{n}_K = (n_k)_0^{K-1}$

Comment estimer  $S_n(f)$  puis  $\mathbf{x}$  ?

# Plan

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Outline

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 **Suppression du bruit de fond**
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Périodogramme

## Définition

C'est la transformée de Fourier sur  $K$  échantillons

$$P_K(f) = \frac{1}{K} \left| \sum_{k=0}^{K-1} n_k e^{-i \frac{2\pi}{K} f k} \right|^2$$

$$\lim_{K \rightarrow +\infty} P_K(f) = S_n(f)$$

# Périodogramme

## Propriété

C'est un estimateur asymptotiquement non biaisé, mais non consistant !

- Biais :

$$E [P_K(f)] = S_n(f) \star \hat{w}_B(e^{i2\pi f})$$

où  $w_B$  est la fenêtre triangulaire (de Bartlett).

$$\hat{w}_B(e^{i2\pi f}) = \frac{1}{N} \left[ \frac{\sin(N\pi f)}{\sin(\pi f)} \right]^2$$

$$\lim_{K \rightarrow +\infty} E [P_K(f)] = S_n(f)$$

- Variance :

$$\text{Var} \{P_K(f)\} = S_n(f)^2$$

# Méthode de Bartlett

**Idée :** faire une moyenne de petits périodogramme.

## Algorithme

- 1 On découpe  $\mathbf{n}_K$  en  $I$  périodogramme de même tailles  $M = K/I$
- 2 On moyenne ces périodogrammes :

$$\hat{P}_M(f) = \sum_{i=1}^I P_M^i(f)$$

# Méthode de Bartlett

## Propriétés

Le périodogramme moyenné est un estimateur consistant

- Biais :

$$\mathbb{E} \left[ \hat{P}_M(f) \right] = S_n(f) \star \hat{w}_B(e^{i2\pi f})$$

- Variance :

$$\text{Var} \left\{ \hat{P}_M(f) \right\} \simeq \frac{1}{M} S_n(f)$$

Mais on perd en résolution par rapport au périodogramme

# Méthode de Welch

Deux modifications à la méthode de Bartlett :

- 1 Les sous-séquences de bruit peuvent se chevaucher
- 2 On utilise le périodogramme modifié : i.e. on multiplie les séquences par une fenêtre (en général de Hamming ou de Hann)

Cette méthode permet de réduire la variance de la méthode de Bartlett

# Plan

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Outline

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 **Suppression du bruit de fond**
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Calcul du filtre optimal

## Rappel du modèle de bruit

$$\mathbf{y} = \mathbf{x} + \mathbf{n}$$

avec  $\mathbf{n}$  bruit gaussien de densité spectrale  $S_n(f)$ .

- On cherche à estimer  $\mathbf{x}$  par un filtre  $\mathbf{h}$  :

$$\hat{\mathbf{x}} = \mathbf{h} \star \mathbf{y}$$

- tel que l'erreur quadratique soit minimale :

$$\mathbf{h} = \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{h} \star \mathbf{y}\|_2^2$$

## Filtre de Wiener

La solution est donnée dans le domaine de Fourier par :

$$h(f) = \frac{\mathbb{E} [|\hat{x}(f)|^2]}{\mathbb{E} [|\hat{x}(f)|^2] + S_n(f)} = \frac{\mathbb{E} [|\hat{x}(f)|^2]}{\mathbb{E} [|\hat{y}(f)|^2]}$$

# Filtre de Wiener : inconvénients

## Filtre de Wiener

La solution est donnée dans le domaine de Fourier par :

$$h(f) = \frac{\mathbb{E} [|\hat{x}(f)|^2]}{\mathbb{E} [|\hat{x}(f)|^2] + S_n(f)} = \frac{\mathbb{E} [|\hat{x}(f)|^2]}{\mathbb{E} [|\hat{y}(f)|^2]}$$

- Il faut estimer  $\mathbb{E} [|\hat{x}(f)|^2]$
- $\mathbf{x}$  ne peut pas être supposé comme stationnaire ! Donc  $\mathbb{E} [|\hat{x}(f)|^2]$  varie au cours du temps.

# Non stationnarité de $\mathbf{x}$

- On procède à un fenêtrage de  $\mathbf{x}$ , puis on calcule la TF dans chaque fenêtre : principe de la STFT (ou Gabor)!

$$\mathbf{y} = \mathbf{x} + \mathbf{n}$$
$$Y(t, f) = X(t, f) + N(t, f), \forall (t, f)$$

- On cherche à estimer  $X(t, f)$  à partir de  $Y(t, f)$  par un estimateur diagonal, i.e.

$$\hat{X}(t, f) = H(t, f)Y(t, f)$$

de sorte que

$$H = \operatorname{argmin}_H \sum_{t, f} |X(t, f) - H(t, f)Y(t, f)|^2$$

# Solution du filtrage à la Wiener

$$\hat{X}(t, f) = H(t, f)Y(t, f)$$

avec

$$H(t, f) = \frac{\mathbb{E}[|X(t, f)|^2]}{\mathbb{E}[|X(t, f)|^2] + S_n^w(f)} = \frac{\mathbb{E}[|X(t, f)|^2]}{\mathbb{E}[|Y(t, f)|^2]}$$

où  $S_n^w(f)$  est l'estimateur de Welch de  $S_n$ , utilisant la même fenêtre et le même overlap que pour calculer la STFT de  $Y$ .

Reste le problème d'estimer  $\mathbb{E}[|X(t, f)|^2]$

# Estimation de $E[|X(t, f)|^2]$

On a

$$Y(t, f) = X(t, f) + N(t, f)$$

$\mathbf{x}$  et  $\mathbf{n}$  étant indépendant, on écrit

$$E[|Y(t, f)|^2] = E[|X(t, f)|^2] + S_{\mathbf{n}}(f)$$

Ce qui donne comme estimateur :

$$E[|X(t, f)|^2] = E[|Y(t, f)|^2] - S_{\mathbf{n}}(f)$$

On a obligatoirement  $E[|X(t, f)|^2] \geq 0$  :

$$E[|X(t, f)|^2] = (E[|Y(t, f)|^2] - S_{\mathbf{n}}(f))^+$$

Ce qui, avec un estimateur ponctuel pour  $E[|Y(t, f)|^2]$  et l'estimateur de Welch pour la dsp du bruit, donne :

$$H(t, f) = \frac{(|Y(t, f)|^2 - S_{\mathbf{n}}^w(f))^+}{|Y(t, f)|^2}$$

# Estimateur diagonal de Wiener empirique

En pratique, on a donc l'estimateur suivant :

Estimateur diagonal de Wiener "empirique"

$$\begin{aligned}\hat{X}(t, f) &= H(t, f)Y(t, f) = \frac{(|Y(t, f)|^2 - S_n^w(f))^+}{|Y(t, f)|^2} Y(t, f) \\ &= Y(t, f) \left(1 - \frac{S_n^w(f)}{|Y(t, f)|^2}\right)^+\end{aligned}$$

En définissant le rapport signal à bruit ponctuel

$$\rho(t, f) = \frac{(|Y(t, f)|^2 - S_n^w(f))^+}{S_n^w(f)}$$

on peut réécrire l'estimateur comme

$$\hat{X}(t, f) = \frac{\rho(t, f)}{1 - \rho(t, f)} Y(t, f)$$

# Plan

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - **Soustraction spectrale**
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Outline

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 **Suppression du bruit de fond**
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Soustraction spectrale de puissance

On repart de

$$E [ |Y(t, f)|^2 ] = E [ |X(t, f)|^2 ] + S_n(f)$$

donnant l'estimation

$$E [ |X(t, f)|^2 ] = E [ |Y(t, f)|^2 ] - S_n(f)$$

On a donc comme estimation de  $|X(t, f)|$  :

$$|\widehat{X}(t, f)| = \sqrt{|Y(t, f)|^2 - S_n(f)}$$

# Soustraction spectrale de puissance

## Soustraction spectrale de puissance

En conservant la phase inchangée par rapport à l'observation et en respectant la positivité, on obtient l'estimateur de soustraction spectrale de puissance :

$$\begin{aligned}\hat{X}(t, f) &= Y(t, f) \sqrt{\frac{(|Y(t, f)|^2 - S_n^w(f))^+}{|Y(t, f)|^2}} \\ &= \sqrt{\frac{\rho(t, f)}{1 - \rho(t, f)}} Y(t, f)\end{aligned}$$

# Inconvénients

- L'estimateur ponctuel de  $E[X(t, f)]$  est un mauvais estimateur
- Apparition de "bruits musicaux"

Comment améliorer l'estimation en pratique ?

# Plan

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Outline

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 **Suppression du bruit de fond**
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Estimateur diagonaux

- On cherche une estimation diagonal de  $X(t, f)$

$$\hat{X}(t, f) = A(t, f)Y(t, f)$$

- On introduit deux paramètres de puissance afin d'écrire les estimateurs précédent sous une forme générale :

$$A(t, f) = \left( \frac{|Y(t, f)|^\alpha - S_n(f)}{|Y(t, f)|^\alpha} \right)^\beta$$

- Ainsi :
  - $\alpha = 2, \beta = 1$  : estimateur de type Wiener
  - $\alpha = 2, \beta = \frac{1}{2}$  : soustraction spectrale de puissance
  - $\alpha = 1, \beta = 1$  : soustraction spectrale d'amplitude

# Estimateurs diagonaux

- On cherche une estimation diagonale de  $X(t, f)$

$$\hat{X}(t, f) = A(t, f)Y(t, f)$$

- On introduit un paramètre  $\lambda$  de sur/sous-estimation du bruit, afin de réduire le bruit musical :

$$A(t, f) = \left( \frac{|Y(t, f)|^\alpha - \lambda S_n(f)}{|Y(t, f)|^\alpha} \right)^\beta$$

# Estimateurs diagonaux

Enfin, on laisse un "seuil" de bruit minimum, qui permettra de "masquer" certains bruits musicaux :

$$A(t, f) = \max \left( \gamma S_n(f), \left( \frac{|Y(t, f)|^\alpha - \lambda S_n(f)}{|Y(t, f)|^\alpha} \right)^\beta \right)$$

avec  $0 < \gamma < 1$ .

# Modèles hybrides

Plusieurs paramètres interviennent :

- Les hyperparamètres  $\gamma$ ,  $\beta$ ,  $\alpha$ ,  $\lambda$ 
  - Les choix :  $\alpha = 2$ ,  $\beta = 1$  donnent de bons résultats. . .
  - $\lambda$  doit être laissé comme paramètre libre
  - $\gamma \ll 1$
- Quelle forme de fenêtre, quelle taille de fenêtre, quel chevauchement ?

# Choix de la fenêtre

## Exemples de fenêtre

- Bartlett (ou triangulaire)

$$w(n) = \begin{cases} \frac{2}{N-1} & 0 \leq n \leq \frac{N-1}{2} \\ 2 - \frac{2n}{N-1} & \frac{N-1}{2} \leq n \leq N-1 \\ 0 & \text{sinon} \end{cases}$$

- Hann

$$w(n) = \begin{cases} 0.5 - 0.5 \cos\left(\frac{2\pi n}{N-1}\right) & 0 \leq n \leq N-1 \\ 0 & \text{sinon} \end{cases}$$

# Choix de la fenêtre : forme

## Exemples de fenêtre

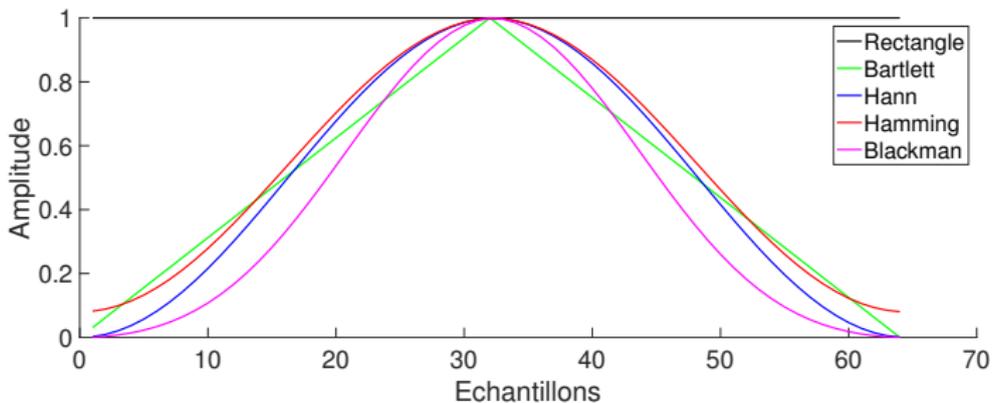
- Hamming

$$w(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) & 0 \leq n \leq N-1 \\ 0 & \text{sinon} \end{cases}$$

- Blackman

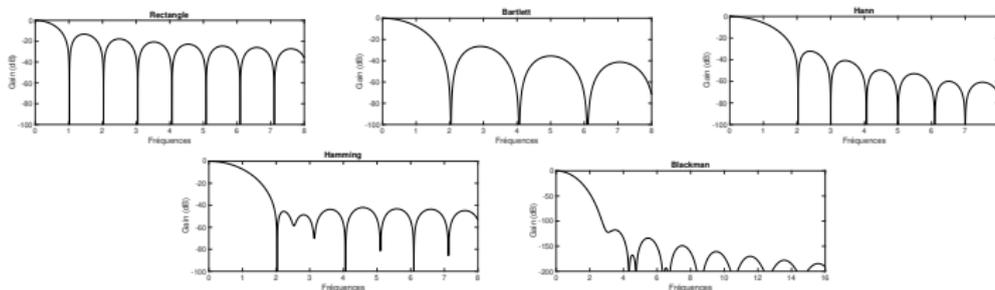
$$w(n) = \begin{cases} 0.42 - 0.5 \cos\left(\frac{2\pi n}{N-1}\right) + 0.08 \cos\left(\frac{4\pi n}{N-1}\right) & 0 \leq n \leq N-1 \\ 0 & \text{sinon} \end{cases}$$

# Choix de la fenêtre : forme



Fenêtres populaires : représentation temporelle

# Choix de la fenêtre : forme



Fenêtres populaires : gain fréquentiel

# Choix de la fenêtre : forme

## Caractéristiques principales

- Largeur du lobe principal
- Rapport d'amplitude entre le lobe principal et le lobe secondaire
- L'atténuation minimale en bande atténuée

Type de fenêtre	Largeur du lobe principal	Rapport d'amplitude	Atténuation minimale
Rectangulaire	$4\pi/N$	-13 dB	-21 dB
Bartlett	$8\pi/N$	-25 dB	-25 dB
Hann	$8\pi/N$	-31 dB	-44 dB
Hamming	$8\pi/N$	-41 dB	-53 dB
Blackman	$12\pi/N$	-57 dB	-74 dB

## Choix de la fenêtre : taille

Le choix de la taille de fenêtre est un compromis sur le principe d'incertitude d'Heisenberg :

- fenêtre courte : les "transitoires" sont privilégiés
- fenêtre longue : les "tonals" sont privilégiés

Bon compromis : 1024 échantillons ( $\simeq 23$  ms) pour un taux d'échantillonnage de 44.1 kHz

## Choix de la fenêtre : overlap

En pratique, on choisit un overlap de

- 50% si la rapidité doit être privilégiée
- 75% pour une amélioration des résultats en pratique (plus de calculs)

# Modèles hybrides

## Définition

$$\begin{aligned}\mathbf{y} &= \mathbf{x} + \mathbf{n} \\ &= \mathbf{x}_{trans} + \mathbf{x}_{ton} + \mathbf{n} \\ &= \mathbf{\Phi}_{trans}\boldsymbol{\alpha}_{trans} + \mathbf{\Phi}_{ton}\boldsymbol{\alpha}_{ton} + \mathbf{n}\end{aligned}$$

(cf. cours approximation parcimonieuse)

- On utilise deux tailles de fenêtre
  - Une adaptée aux composantes transitoires (64 à 256 échantillons, ie 1.5 ms à 6 ms)
  - Une adaptée aux composantes tonales (1024 à 4096 échantillons, ie 23 ms à 93 ms)

# Modèles Hybrides

## Algorithme pratique

- 1 Débruitage avec une fenêtre adaptée au tonal : on obtient  $\hat{\mathbf{x}}_{ton}$
- 2 Débruitage avec une fenêtre adaptée au transitoire sur le résidu  $\mathbf{y} - \hat{\mathbf{x}}_{ton}$  : on obtient  $\hat{\mathbf{x}}_{trans}$
- 3  $\hat{\mathbf{x}} = \hat{\mathbf{x}}_{trans} + \hat{\mathbf{x}}_{ton}$

# Modèles non diagonaux

On peut utiliser tous les opérateurs de seuillage vus dans le cours "approximations parcimonieuses" :

- Seuillage dur (estimateur diagonal)
- WG-Lasso
- Persistent Empirical Wiener

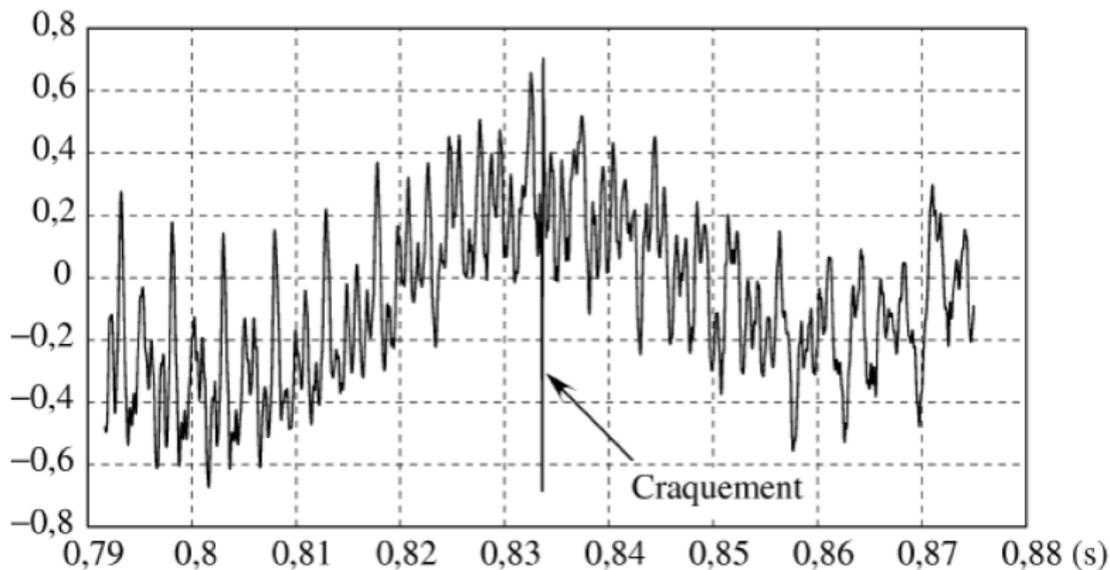
# Plan

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Outline

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 **Suppression des clics/craquements**
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Clicks/craquements



# Hypothèses

- Les clics/craquements sont rares (1% des échantillons)
- Leur durée est courte (de 1 à  $\simeq 10$  échantillons)

# Modèles

## Modèle de bruit

$$y_t = x_t + n_t = x_t + i_t v_t$$

- $v_t$  est un processus aléatoire
- $i_t = 0/1$  un processus aléatoire binaire :  $i_t = 1$  lorsque le bruit est présent.

Différence fondamentale avec le bruit de fond : les bruits sont impulsionnels et donc très localisés ! On va travailler directement sur les échantillons temporels.

Quel modèle de signal permet de détecter les clics/cracks ?

# Modèles

## Modèle AR du signal

On suppose que  $\mathbf{x}$  est un signal auto-régressif (AR) d'ordre  $p$  :

$$x_t + \sum_{k=1}^p a_k x_{t-k} = e_t$$

où

- $\mathbf{a} = (a_k)$  sont les paramètres du modèle AR
- $\mathbf{e} \sim \mathcal{N}(0, \sigma_e^2)$  est un bruit blanc gaussien

# Idée générale

$\mathbf{y}$  est donc un processus AR dont quelques échantillons sont corrompus. Si l'on suppose les paramètres du modèle connus (les  $(a_k)$  et  $\sigma_e$ ), le schéma de détection serait :

- 1 Calcul de l'erreur de prédiction :  $\epsilon_t = y_t + \sum_{k=1}^p a_k y_{t-k}$
- 2 Si  $|\epsilon_t| > K\sigma_e$ , alors  $i_t = 1$ , et 0 sinon.

avec  $K \simeq 3$  si l'on suit la règle des  $3\sigma$  pour la détection d'évènements rares.

# Hypothèses

- Les clics sont rares (1% des échantillons)
- Leur durée est courte (de 1 à  $\simeq 10$  échantillons)
- On peut estimer le modèle AR dans une fenêtre courte (quelques centaines d'échantillons)
- En pratique  $20 < p < 50$ .

# Estimation du modèle

- 1 On découpe  $\mathbf{y}$  par fenêtrage de taille  $M$
- 2 On estime les paramètres AR ( $a_k$ ) et  $\sigma_e$  dans chaque frame
- 3 On applique un seuil de détection

# Détection des clics

Dans chaque frame de taille  $M$

- 1 Estimation de la fonction d'autocorrélation

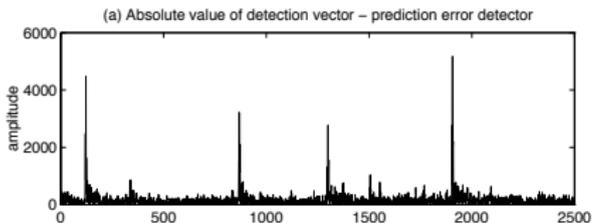
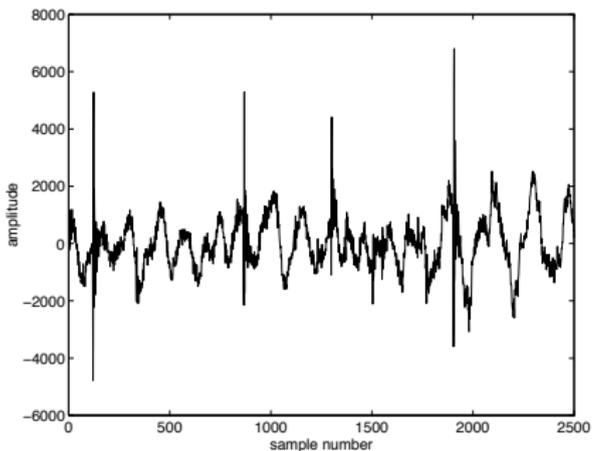
$$\hat{r}(t) = \frac{1}{M} \sum_{k=t+1}^M x_k x_{k-t}$$

- 2 Écriture des équations de Yule-Walker

$$\begin{pmatrix} \hat{r}(0) & \hat{r}(1) & \dots & \hat{r}(p-1) \\ \hat{r}(1) & \hat{r}(0) & \dots & \hat{r}(p-2) \\ & \vdots & & \\ \hat{r}(p-1) & \hat{r}(p-2) & \dots & \hat{r}(0) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{p-1} \end{pmatrix} = \begin{pmatrix} \hat{r}(1) \\ \hat{r}(2) \\ \vdots \\ \hat{r}(p) \end{pmatrix}$$

- 3 Résolution des équations de Yule-Walker par un algorithme de Levinson-Durbin
- 4 Estimation de  $\sigma_e$  :  $\hat{\sigma}_e = r(0) - \hat{\mathbf{a}}^T \mathbf{r}$
- 5 Décision :  $i_t = 1$  si  $(\mathbf{y} + \mathbf{a} \star \mathbf{y})_t > \lambda \hat{\sigma}_e$

# Détection des clics/craquements



# Détection des clics/craquements

Comment reconstruire la partie de signal corrompue ?

# Plan

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 Suppression des clics/craquements
  - Détection des clics/craquements
  - Reconstruction des échantillons

# Outline

- 1 Introduction
- 2 Modèles de bruits
  - Les bruits de fond
  - Bruits "impulsionnels"
- 3 Mesures de qualité
- 4 Suppression du bruit de fond
  - Estimation de la densité spectrale de puissance
  - Filtrage de Wiener
  - Soustraction spectrale
  - Estimateurs diagonaux
- 5 **Suppression des clics/craquements**
  - **Détection des clics/craquements**
  - **Reconstruction des échantillons**

# Reconstruction des échantillons

- Les clics sont détectés
- Le modèle impulsionnel est trop restrictif : en pratique, on considère  $l$  échantillons avant et après les clics/craquements détectés par  $i_t = 1$ .
- Idée général : interpoler le signal en utilisant les échantillons non corrompu avant et après la zone du craquement.

# Interpolation Least Square AR (LSAR)

- Le modèle AR peut s'écrire sous forme matricielle :

$$\mathbf{Ax} = \mathbf{e}$$

avec

$$\mathbf{A} = \begin{pmatrix} a_p & a_{p-1} & \dots & a_1 & 1 & 0 & \dots & \dots & 0 \\ 0 & a_p & a_{p-1} & \dots & a_1 & 1 & 0 & \dots & 0 \\ \vdots & & & & & & & & \\ 0 & \dots & \dots & 0 & a_p & a_{p-1} & \dots & a_2 & 1 \end{pmatrix}$$

- On réordonne  $\mathbf{x}$ , de manière à identifier les échantillons non corrompu et les échantillons corrompu :

$$\mathbf{x} = \mathbf{Kx}_r + \mathbf{Ux}_c$$

où  $\mathbf{U}$  et  $\mathbf{K}$  sont des matrices de réarrangement, formant une partition en colonne de la matrice identité.

# Interpolation Least Square AR (LSAR)

- Le modèle s'écrit

$$\begin{aligned}\mathbf{e} &= \mathbf{A}\mathbf{K}\mathbf{x}_r + \mathbf{A}\mathbf{U}\mathbf{x}_c \\ &= \mathbf{A}_r\mathbf{x}_r + \mathbf{A}_c\mathbf{x}_c\end{aligned}$$

- L'erreur quadratique du modèle de prédiction s'écrit alors :

$$\|\mathbf{e}\|^2 = \mathbf{e}^T \mathbf{e} = (\mathbf{A}_r\mathbf{x}_r + \mathbf{A}_c\mathbf{x}_c)^T (\mathbf{A}_r\mathbf{x}_r + \mathbf{A}_c\mathbf{x}_c)$$

- Pour minimiser cette erreur par rapport à  $\mathbf{x}_c$ , on annule la dérivée

$$\frac{\partial \|\mathbf{e}\|^2}{\partial \mathbf{x}_c} = 2\mathbf{e}^T \frac{\partial \mathbf{e}}{\partial \mathbf{x}_c} \quad (1)$$

$$= 2(\mathbf{A}_r\mathbf{x}_r + \mathbf{A}_c\mathbf{x}_c)^T \mathbf{A}_c = 0 \quad (2)$$

- Les échantillons sont donc reconstruit par

$$\mathbf{x}_c = -(\mathbf{A}_c^T \mathbf{A}_c)^{-1} \mathbf{A}_c^T \mathbf{A}_r \mathbf{x}_r$$