



Introduction à l'Audio Numérique

Cours 2 et 3 : Analyse du signal

Julien Pinquier



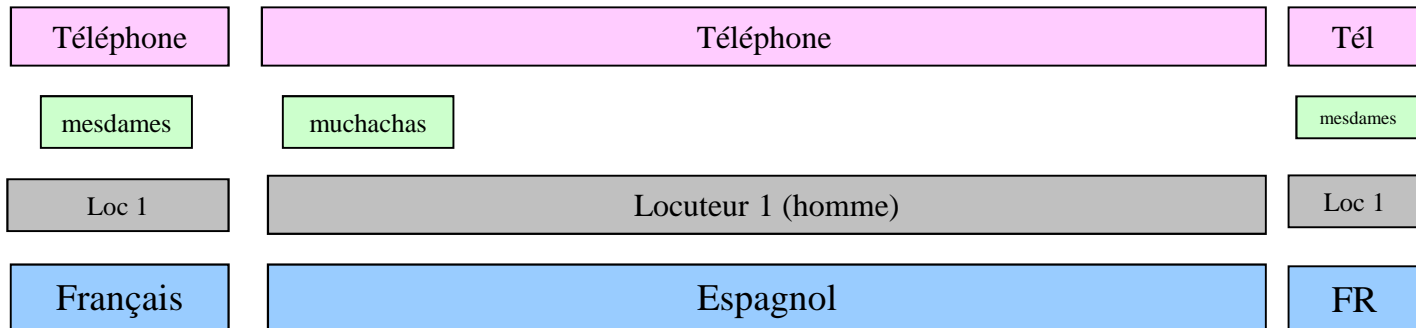
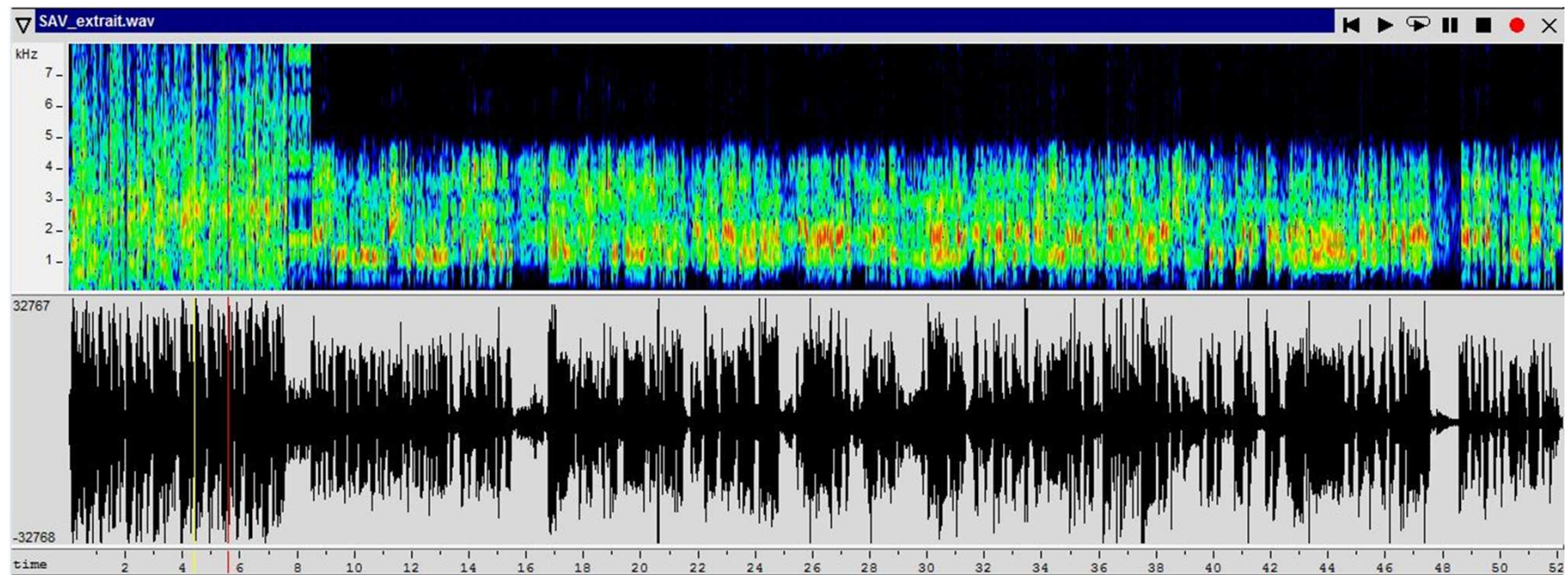
Cours 2 et 3 : plan



- ✧ Introduction
- ✧ Analyse temporelle
- ✧ Analyse fréquentielle
- ✧ Le cepstre
- ✧ Autres paramètres



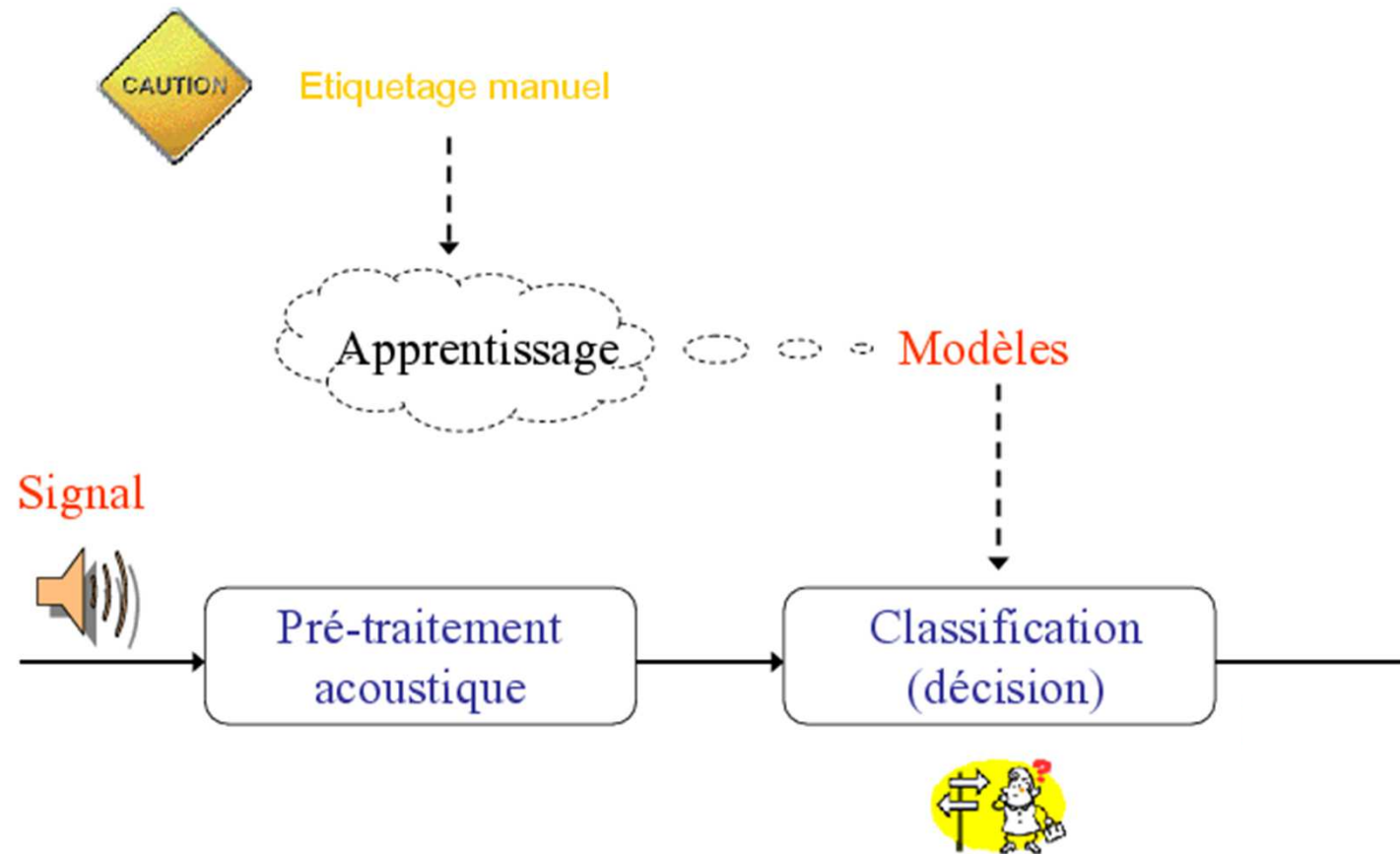
Introduction : que faire avec le son ?



Introduction



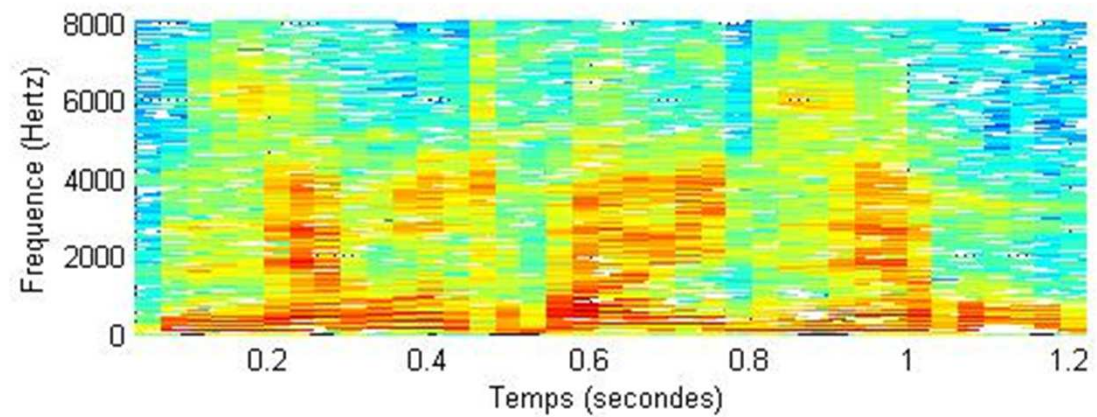
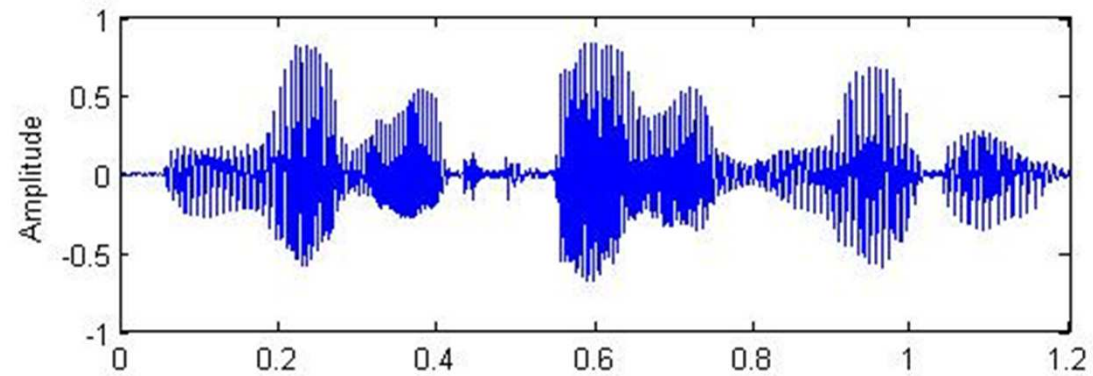
✧ Système classique de reconnaissance des formes



Introduction



✧ Rappel : signal de parole



Introduction



✧ Pourquoi effectuer une paramétrisation ?

- ◆ zone voisée / non voisée
- ◆ fréquence fondamentale (F_0)
- ◆ spectre (pour passer par les fréquences)





✧ Comment effectuer la paramétrisation ?

- ◆ travailler sur des portions de signal et pas sur l'ensemble
- ◆ statistiques à court terme / statistiques à long terme
- ◆ fenêtres temporelles de 10 à 30 ms / plusieurs secondes

- ◆ Spectre
 - Transformée de Fourier à court terme
 - TF Discrète (algorithme TF Rapide)

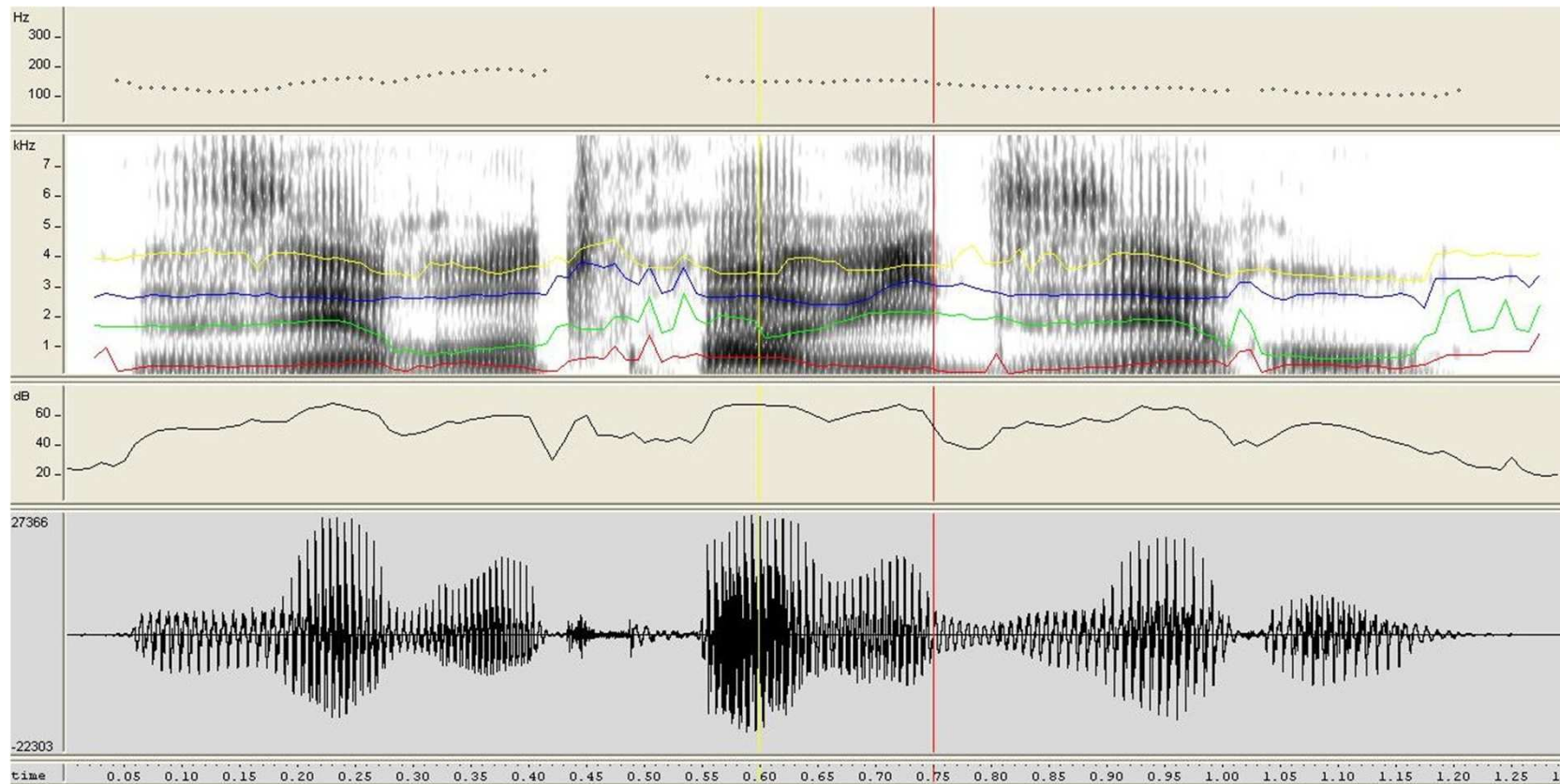
- ◆ Spectrogramme



Introduction



✧ Avec un logiciel classique « wavesurfer »



Introduction



- ✧ Analyse sur une fenêtre glissante
 - ◆ Ordre de 20 ms
 - ◆ Recouvrement (moitié en général)



Introduction



✧ Quelques paramètres

- ◆ Temporels : Energie, ZCR
- ◆ Fréquentiels : issus de la DSP, coefficients spectraux
- ◆ Autre : Mel Frequency Cepstrum Coefficients (MFCC)
- ◆ Etc.





✧ **Exercice** : soit un enregistrement sonore stéréo, échantillonné à une fréquence de 16 kHz. L'analyse se fait sur des fenêtres de 256 points (avec recouvrement sur la moitié).

1. Quel est le nombre de points capturés par intervalle de temps ?
2. Quelle est la durée (en ms) d'une fenêtre d'analyse ?
3. Quel est le nombre de fenêtres (vecteurs) à traiter pour un fichier de 10 minutes ?





✧ Exercice (solution) :

1. Quel est le nombre de points capturés par intervalle de temps ?
2. Quelle est la durée (en ms) d'une fenêtre d'analyse ?
3. Quel est le nombre de fenêtres (vecteurs) à traiter pour un fichier de 10 minutes ?



Cours 2 et 3 : plan



- ✧ Introduction
- ✧ Analyse temporelle
- ✧ Analyse fréquentielle
- ✧ Le cepstre
- ✧ Autres paramètres

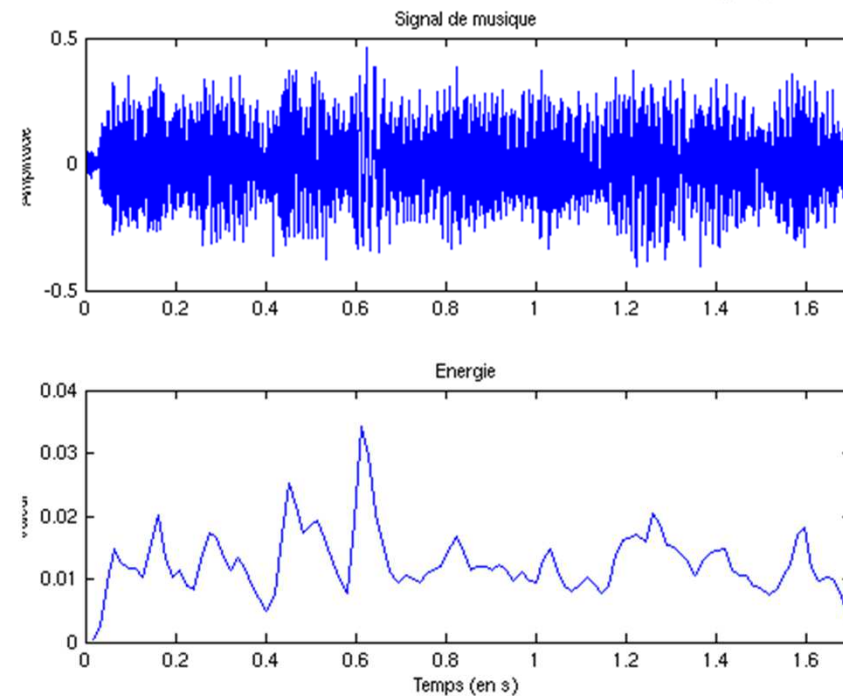
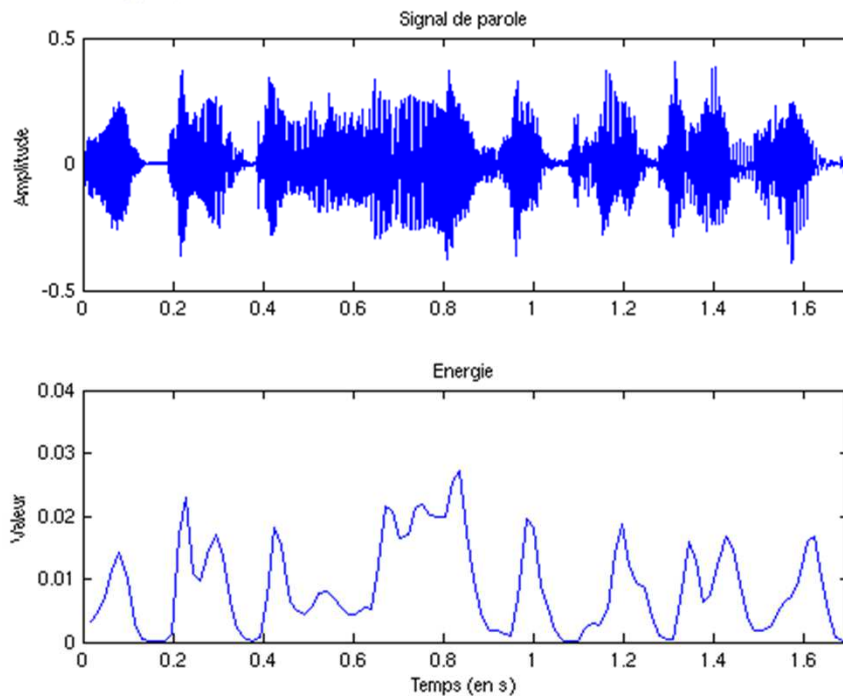


Analyse temporelle



✧ Energie :

- ◆ Discrimination des sons voisés / non-voisés
- ◆ Détection de silence
- ◆ Aide à la décomposition en parole / musique

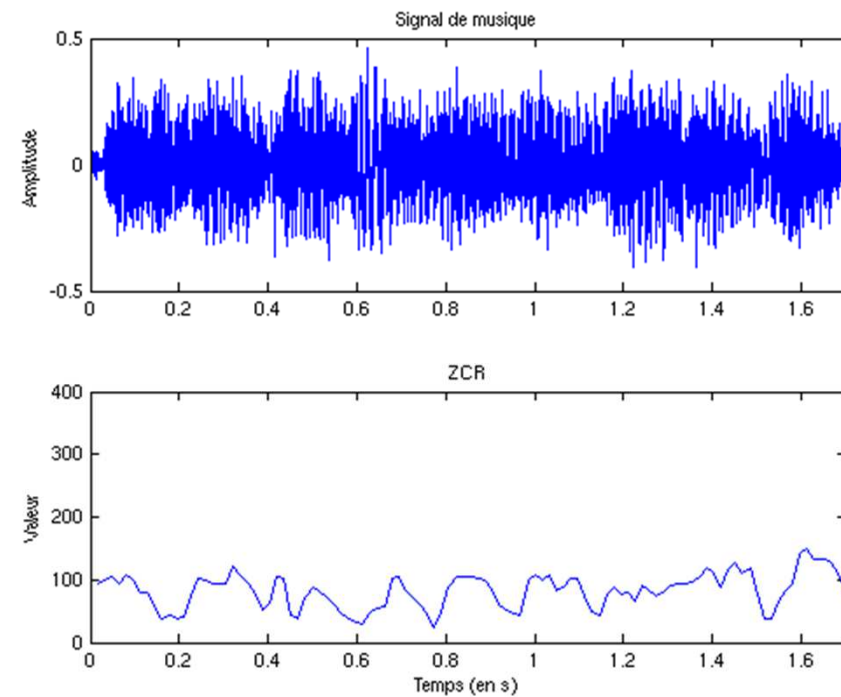
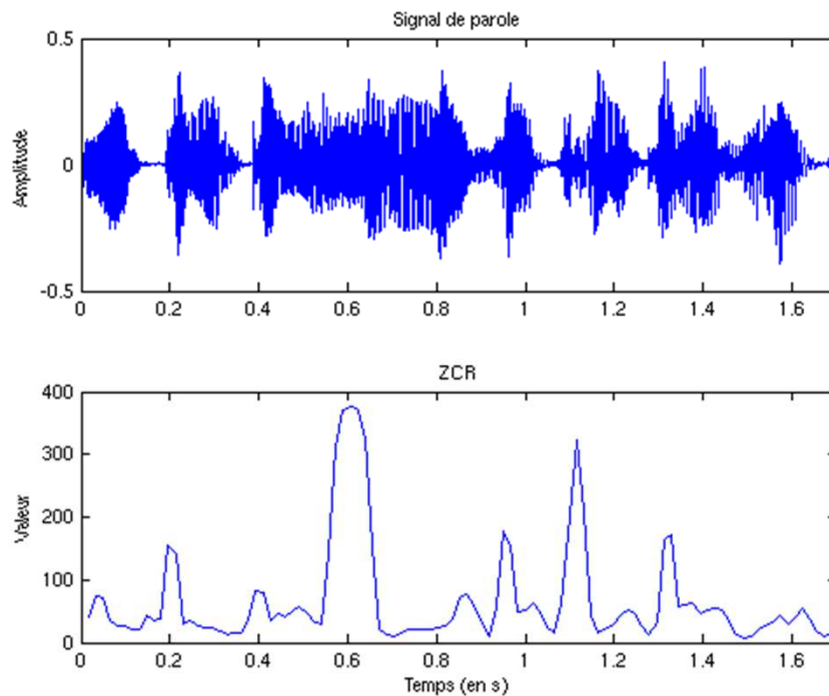


Analyse temporelle



✧ ZCR :

◆ Plus discriminant que l'énergie



Cours 2 et 3 : plan



- ✧ Introduction
- ✧ Analyse temporelle
- ✧ Analyse fréquentielle
- ✧ Le cepstre
- ✧ Autres paramètres





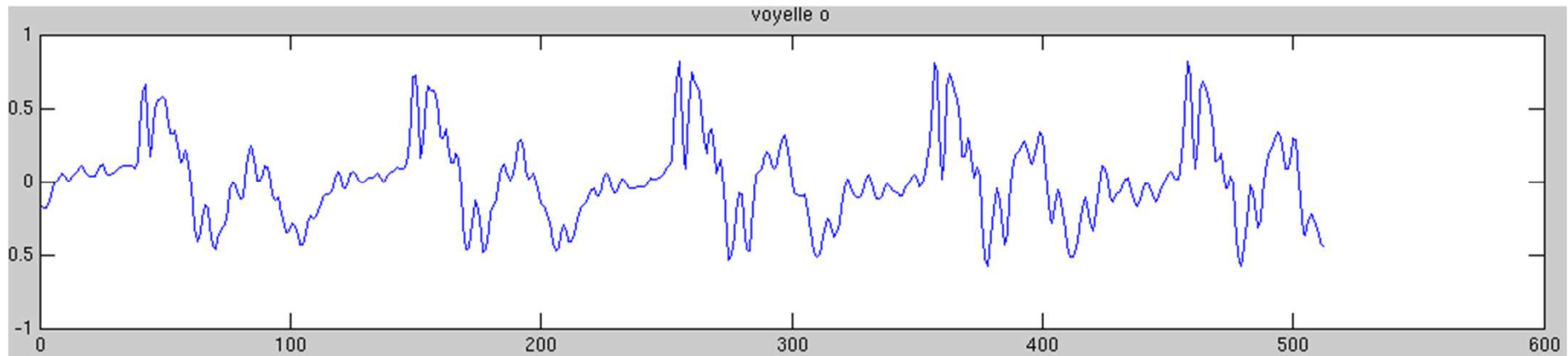
- ✧ Passer dans le domaine fréquentiel : pourquoi et comment ?
- ✧ Fonctionnement du système auditif
 - ◆ l'oreille effectue une analyse spectrale de l'onde acoustique reçue
 ↔ perçoit la composition fréquentielle des sons (=spectre)



Analyse fréquentielle



✧ Spectre





✧ Spectre



◆ Signal x_n





✧ Spectre



- ◆ Signal x_n
- ◆ Préaccentuation des aigus





- ✧ Fenêtrage : pourquoi ?
 - ◆ Signal quasi-stationnaire sur fenêtres de 10 à 30 ms
 - ◆ Pour limiter les effets de bord (réduire les discontinuités)

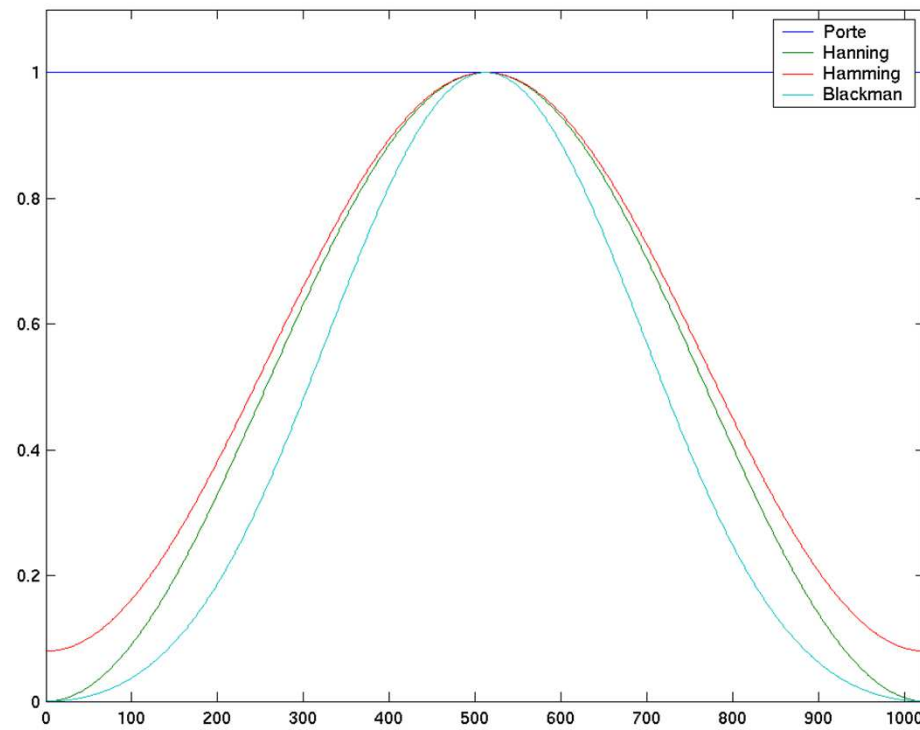
- ✧ Analyse sur une fenêtre glissante

- ✧ Exemples de fenêtres





✧ Fenêtre de Hamming



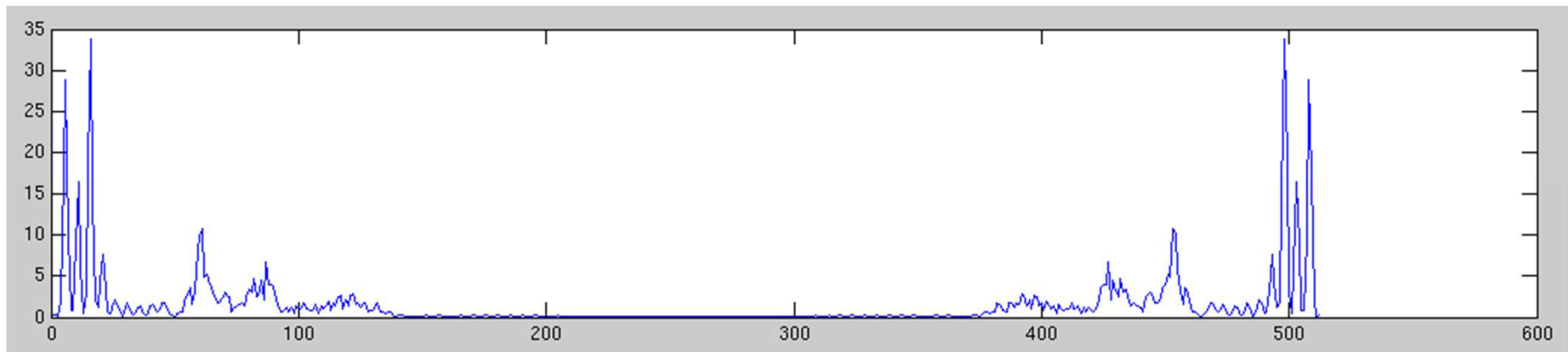
Analyse fréquentielle



✧ Spectre



- ◆ Signal x_n
- ◆ Préaccentuation des aigus
- ◆ Fenêtrage de Hamming

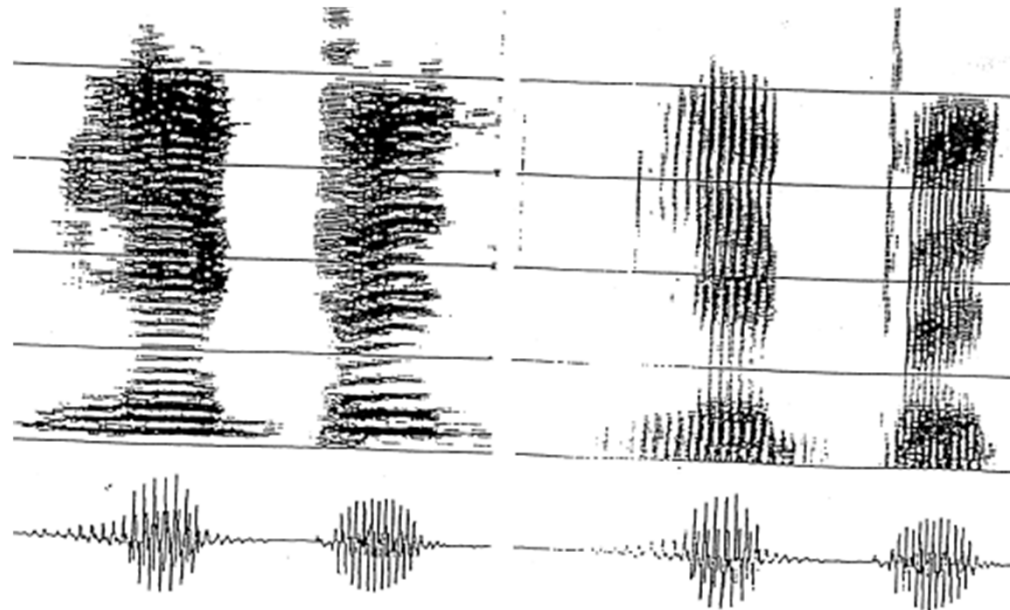




✧ Impact de la longueur de la fenêtre d'analyse

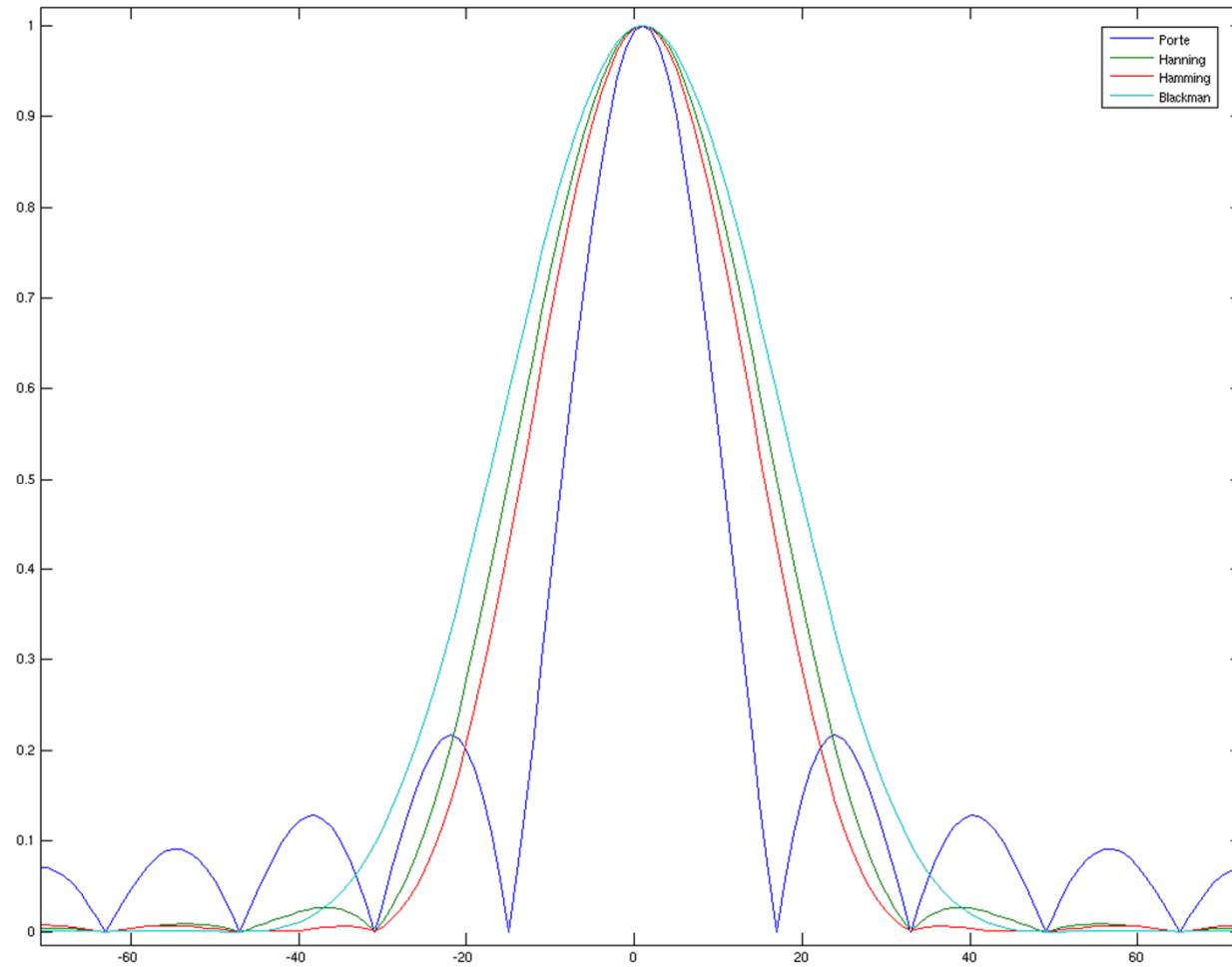
◆ Bande étroite :

◆ Bande large :





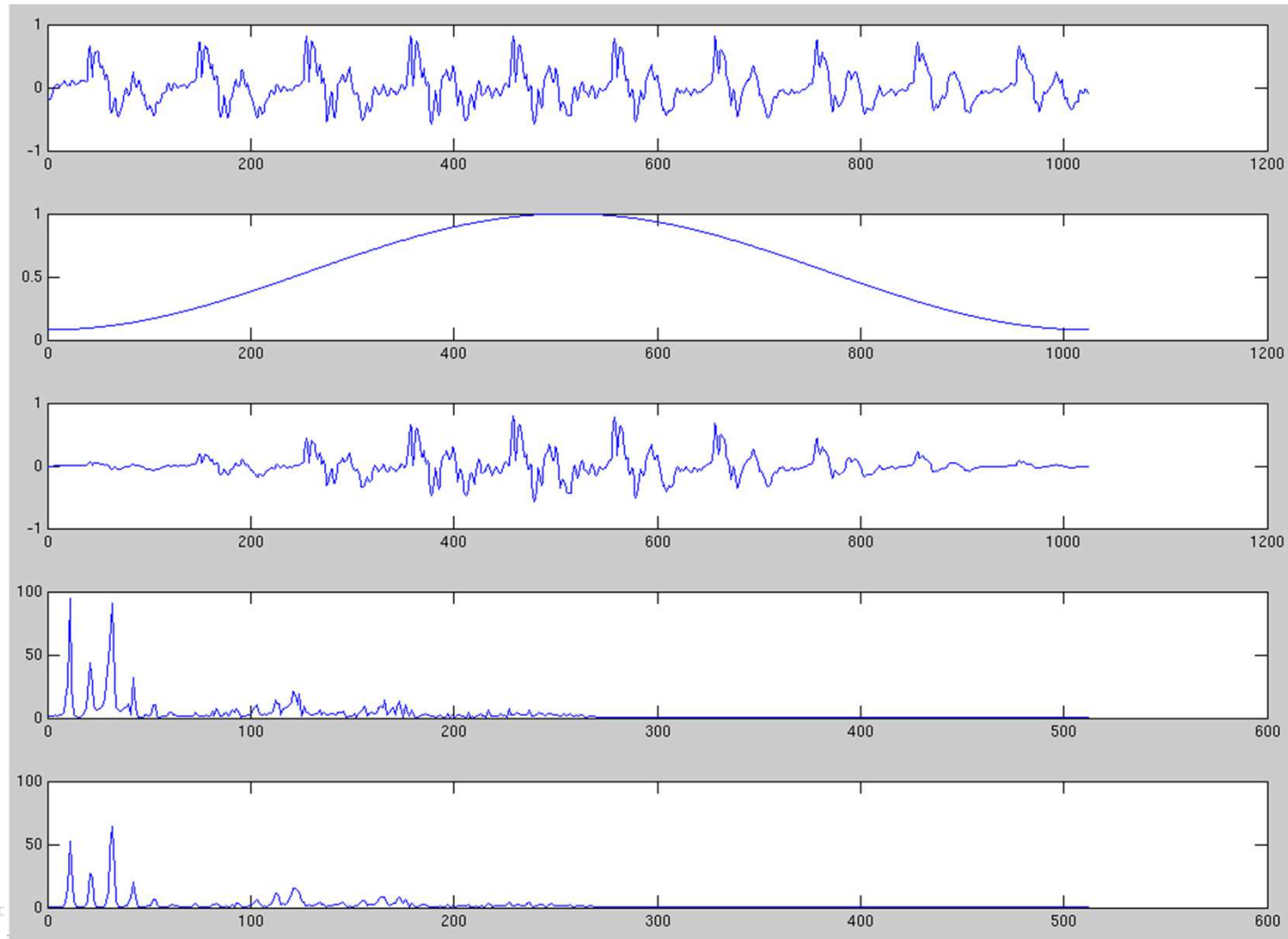
✧ Impact du type de fenêtrage



Analyse fréquentielle



✧ Impact de la fenêtre de Hamming sur un spectre





✧ Utilisation de la DSP (Densité Spectrale de Puissance)

- ◆ Transformée de Fourier de la fonction d'autocorrélation
- ◆ Paramètres de base

Centroïde spectral (centre de gravité)

Flux spectral (variation du spectre)

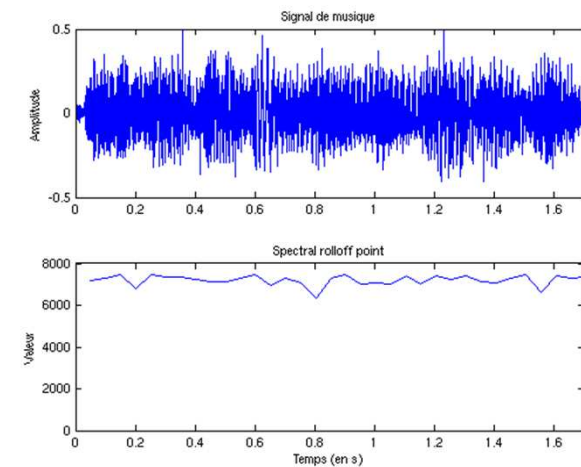
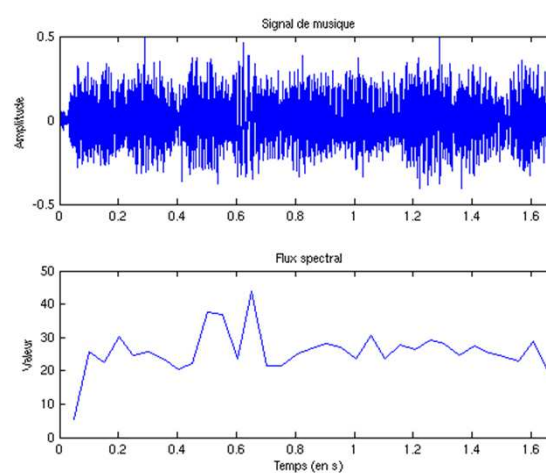
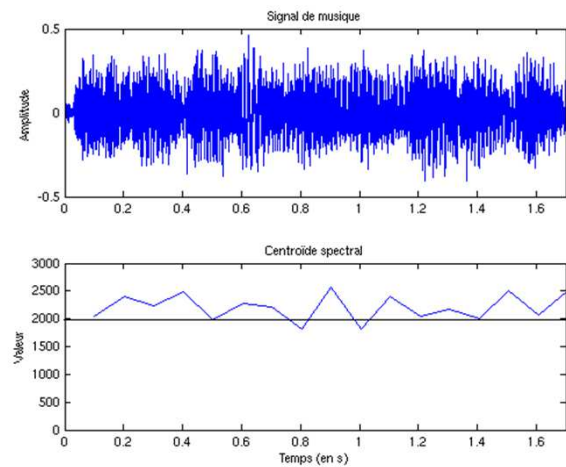
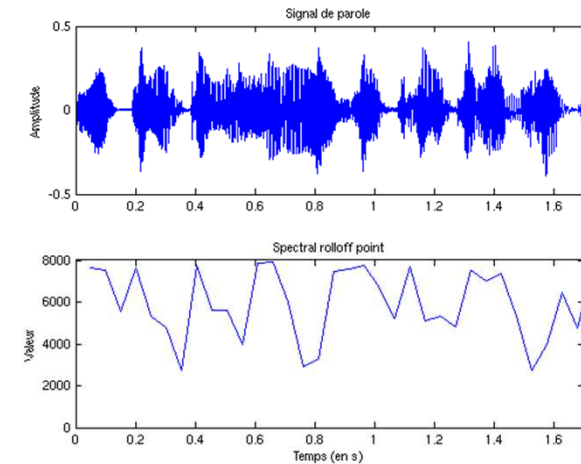
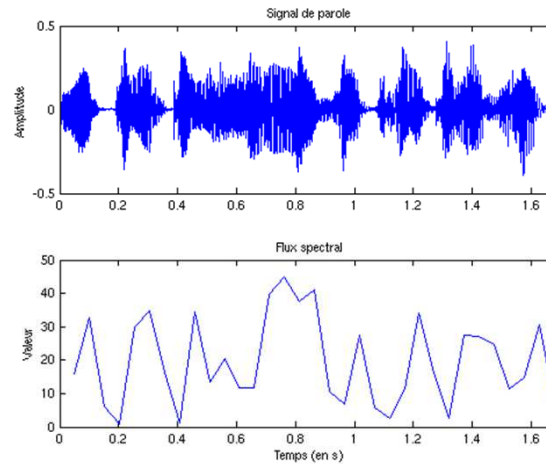
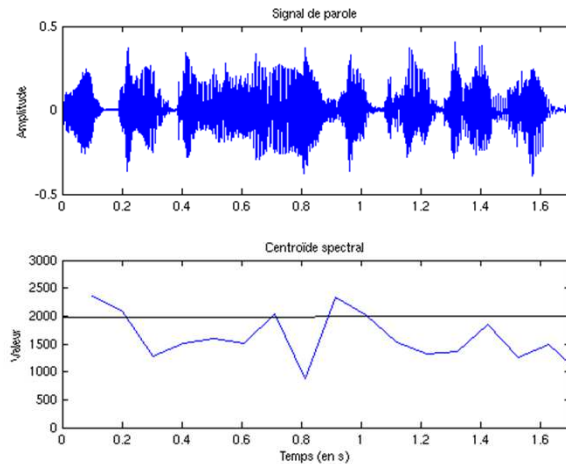
Spectral rolloff point (fréquence de coupure à 95%)



Analyse fréquentielle

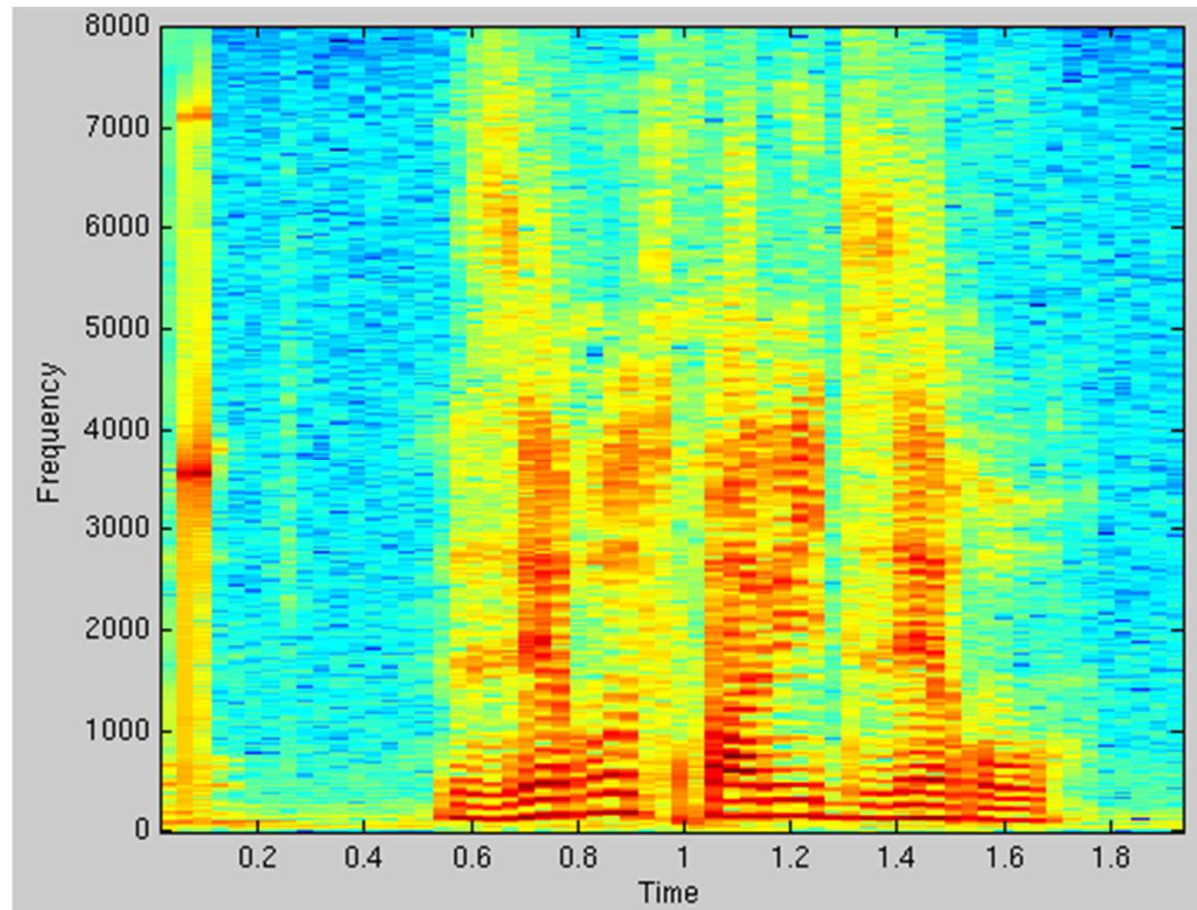


✧ Application : décomposition parole / musique





✧ Spectrogramme



Cours 2 et 3 : plan



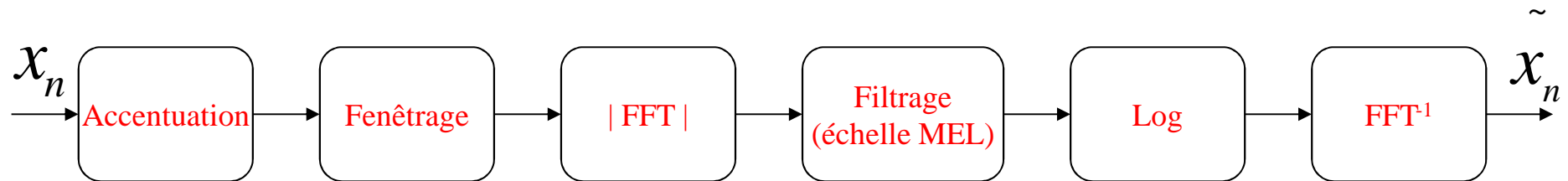
- ✧ Introduction
- ✧ Analyse temporelle
- ✧ Analyse fréquentielle
- ✧ Le cepstre
- ✧ Autres paramètres



Le cepstre



✧ MFCC



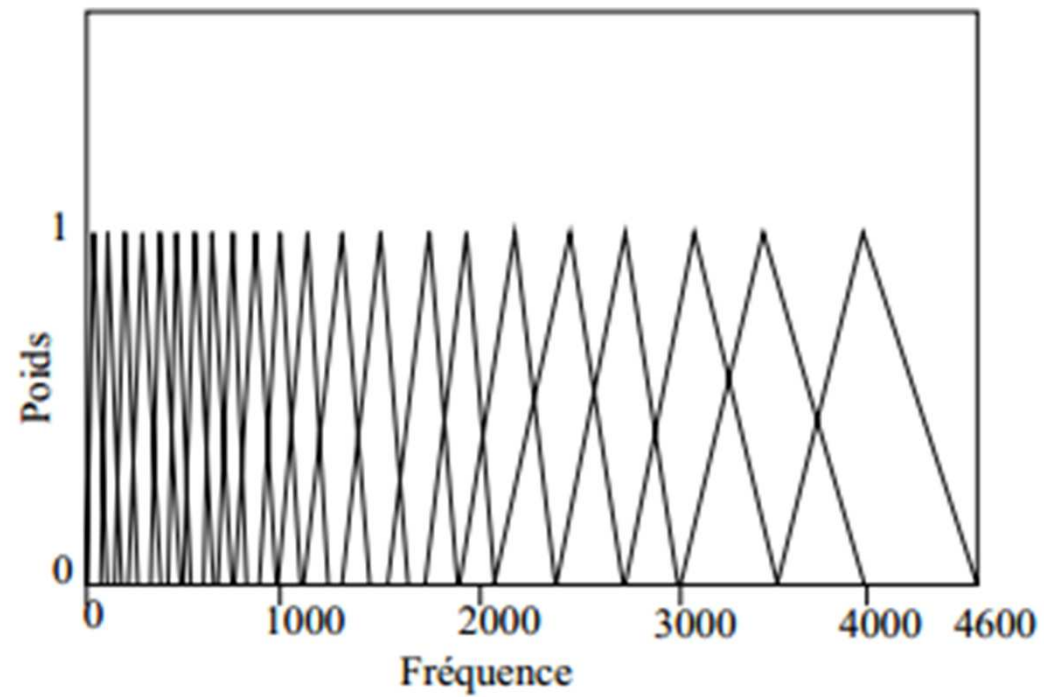
- ◆ Préaccentuation des aigus
- ◆ Fenêtrage de Hamming
- ◆ Transformée de Fourier Rapide



Le cepstre



✧ Echelle perceptive



Le cepstre



✧ Echelles perceptives

- ◆ **Echelle Mel** : linéaire en basse fréquence, logarithmique en haute fréquence

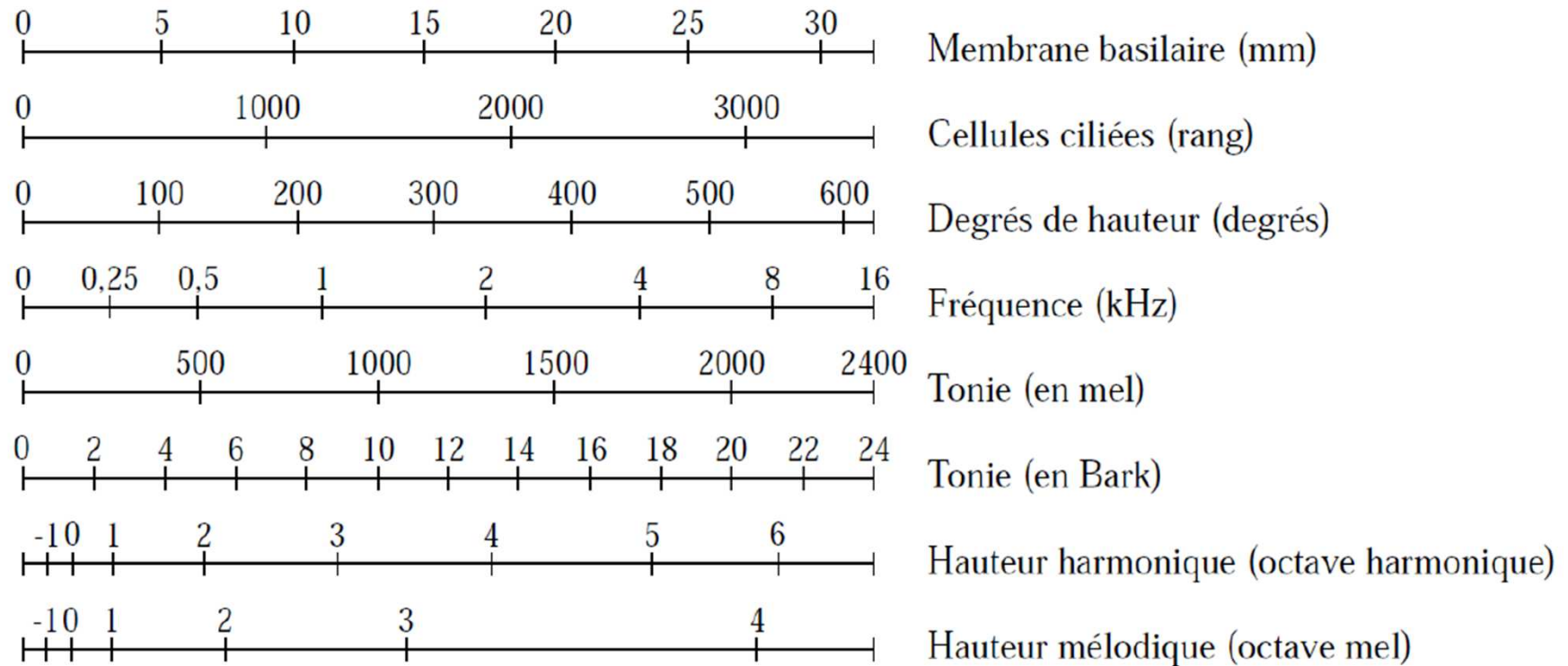
- ◆ **Echelle Bark** :



Le cepstre



✧ Illustration des échelles perceptives



source : Zwicker 1981



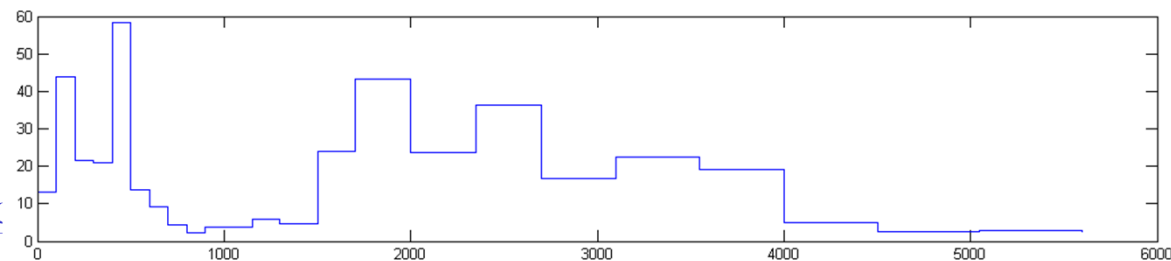
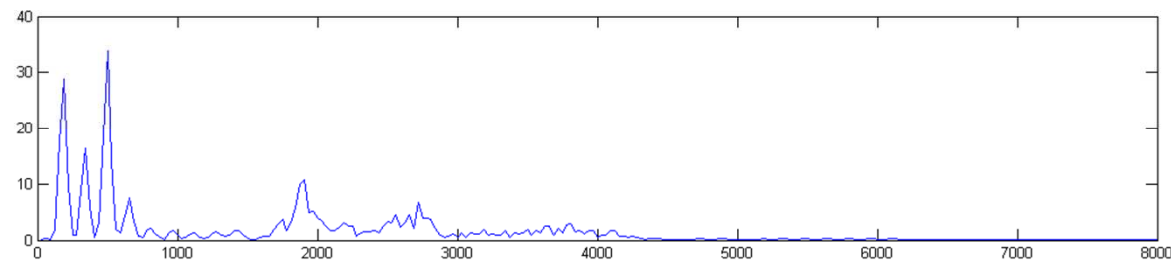
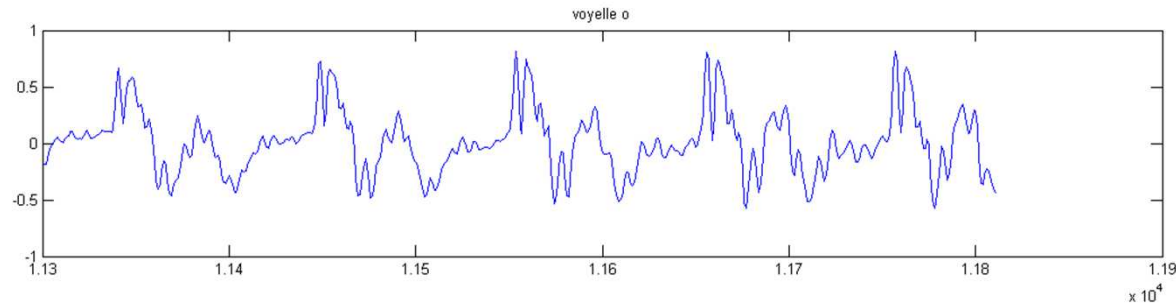
Le cepstre



✧ Illustration de l'échelle MEL

◆ Fréquences de coupure de la pondération triangulaire :

0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1150, 1300, 1500, 1700, 2000, 2350, 2700, 3100, 3550, 4000, 4500, 5050, 5600, 6200, 6850, 7500





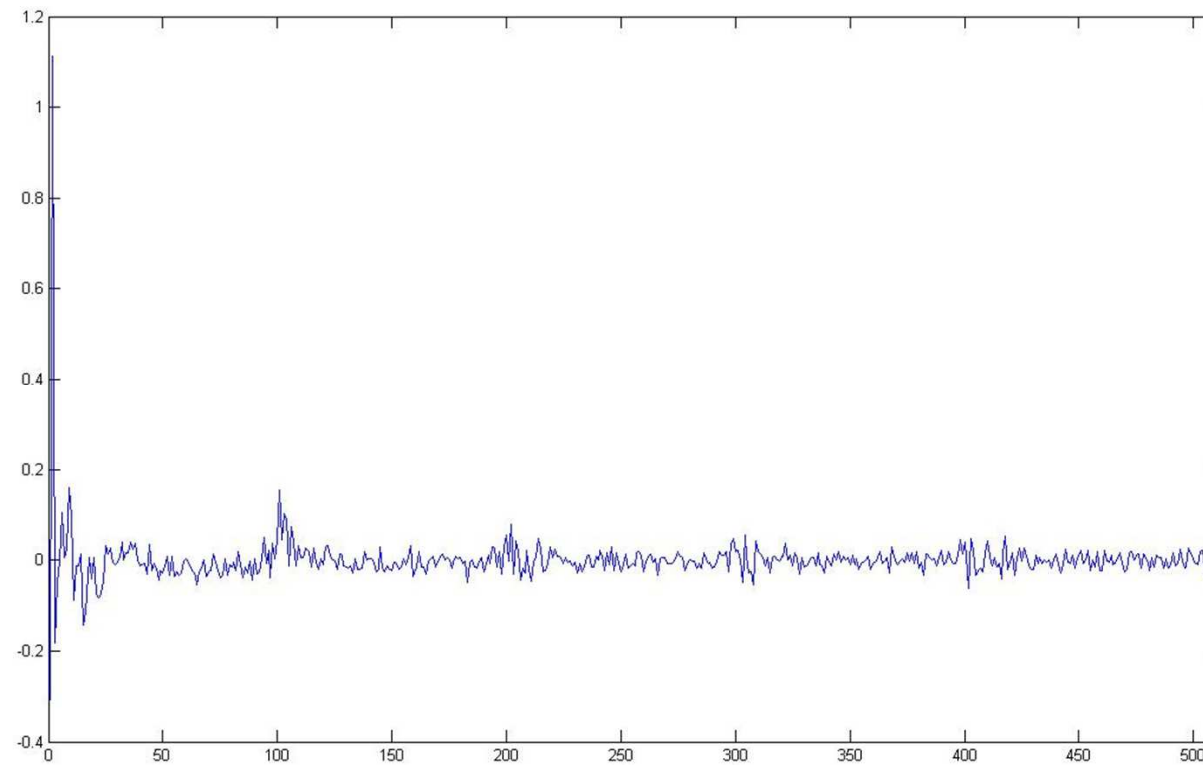
✧ MFCC

- ◆ Déconvolution source/conduit = transformation homomorphique
 - séparer l'excitation du conduit vocal
 - supprimer la fréquence fondamentale (F_0)

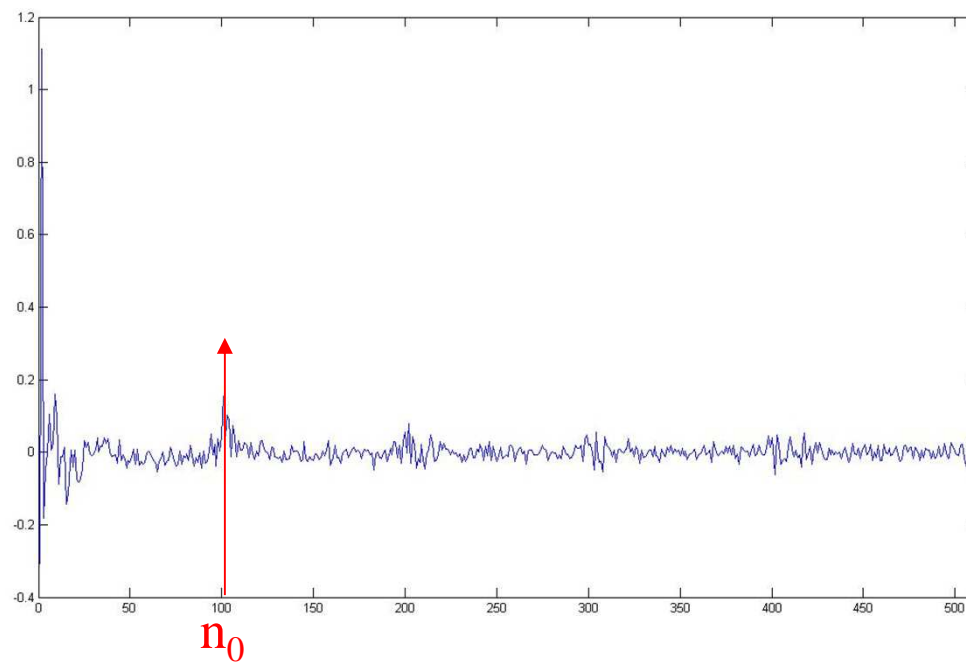




✧ Représentation du cepstre



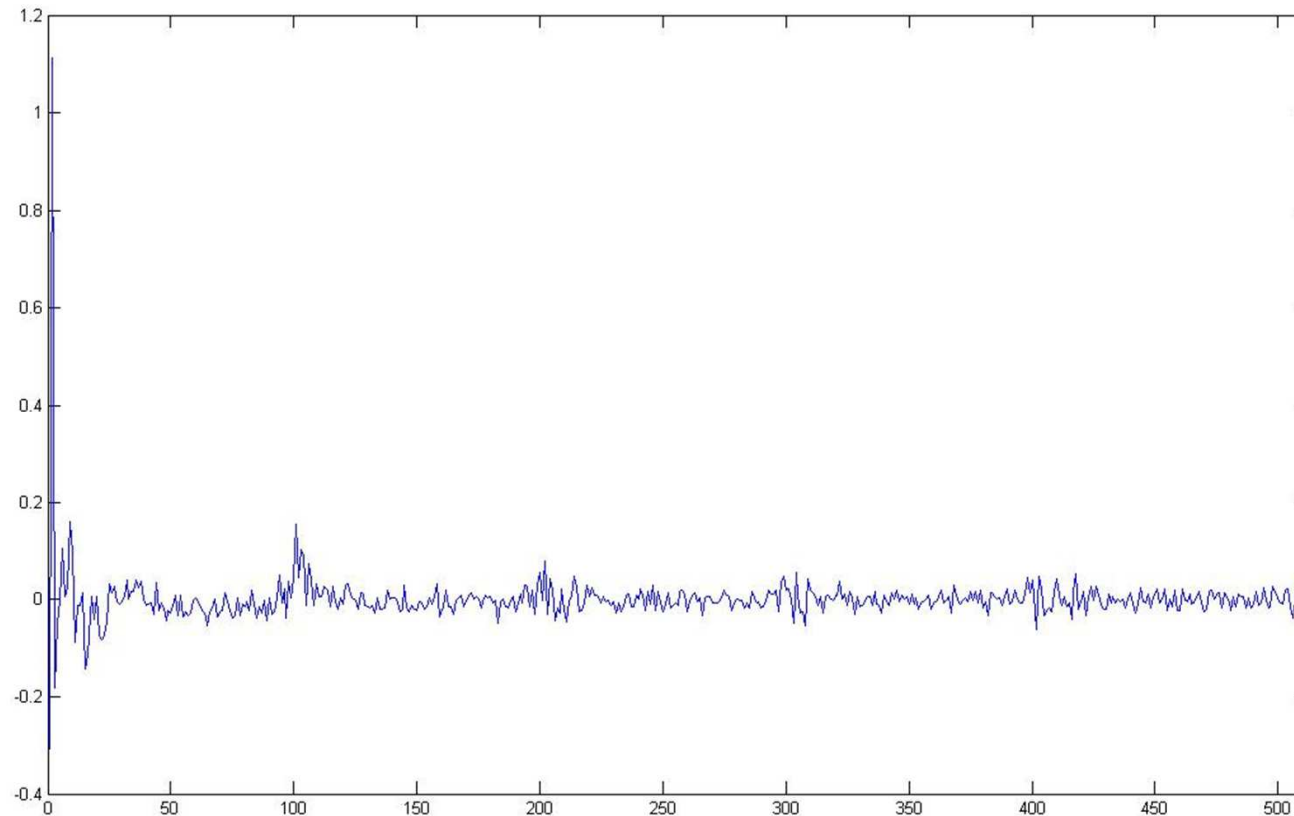
Le cepstre





✧ Exercice

- ◆ Calculer la F_0 du cepstre ci-dessous sachant que $F_e = 16$ kHz et que la FFT a été réalisée sur 1024 points (fenêtre d'analyse de 1024 points)
- ◆ Homme ou femme ?
- ◆ Que vaudrait n_0 pour un enfant ($F_0 = 450$ Hz) ?





✧ Exercice : solution

- ◆ Calculer la F_0 de l'extrait ci-dessous sachant que $F_e = 16$ kHz et que la FFT a été réalisée sur 1024 points (fenêtre d'analyse de 1024 points)
- ◆ Homme ou femme ?
- ◆ Que vaudrait n_0 pour un enfant ($F_0 = 450$ Hz) ?



Cours 2 et 3 : plan



- ✧ Introduction
- ✧ Analyse temporelle
- ✧ Analyse fréquentielle
- ✧ Le cepstre
- ✧ Autres paramètres



Autres paramètres



- ✧ Introduction : spécificités de la musique
- ✧ Fréquence fondamentale
- ✧ Constant-Q Transform
- ✧ Chromas



Autres paramètres : introduction



✧ Spécificités de la musique : rappels

- ◆ Gamme
- ◆ Octave
- ◆ Quinte
- ◆ Quarte
- ◆ L'étalon ➔ le « LA3 » : 440 Hz



Autres paramètres : introduction

✧ Spécificités de la musique : rappels

- ◆ Oreille humaine
- ◆ Gamme tempérée

◆ Rapports de fréquences des notes

do	do#	ré	ré#	mi	fa	fa#	sol	sol#	la	la#	si	do
1	$2^{1/12}$	$2^{1/6}$	$2^{1/4}$	$2^{1/3}$	$2^{5/12}$	$\sqrt{2}$	$2^{7/12}$	$2^{2/3}$	$2^{3/4}$	$2^{5/6}$	$2^{11/12}$	2

Gamme tempérée					
Rapport d'une note à l'autre				R	
Équivalent du physicien				P	
Note	Fréquence				
do ⁷			4186,03		2 ¹²
do ⁶			2093,02		2 ¹¹
do ⁵			1046,51		2 ¹⁰
	ré ⁴		587,33		
do ⁴			523,25		2 ⁹
	si ³		493,88		1,0595
		si b ³	466,16		1,0595
	la ³		440,00		1,0594
		la b ³	415,31		1,0595
	sol ³		392,00		1,0595
		fa # ³	369,99		1,0594
	fa ³		349,23		1,0595
	mi ³		329,63		1,0595
		mi b ³	311,13		1,0595
	ré ³		293,66		1,0595
		do # ³	277,18		1,0594
do ³			261,63		2 ⁸
do ²			130,81		2 ⁷
do ¹			65,41		2 ⁶
do ⁰			32,70		2 ⁵
do ⁻¹			16,35		2 ⁴
do ⁻²			8,15		2 ³

◆ Nomenclature

Diagram of a piano keyboard showing the layout of white and black keys. The white keys are labeled with their names: Do, Ré, Mi, Fa, Sol, La, Si, and Do. The black keys are grouped in pairs and groups of three between the white keys.

- ◆ En musique, deux dimensions :

The diagram illustrates the relationship between chroma and height in music theory. It shows three stacked ellipses representing chroma space, with a vertical axis labeled 'HEIGHT'. The ellipses are labeled with notes: A, B, C, D, E, F, G, A♯, B♯, C♯, D♯, E♯, F♯, G♯. The vertical axis is labeled 'HEIGHT' and the horizontal axis is labeled 'CHROMA'.

Autres paramètres : introduction



- ✧ Spécificités de la musique : lien entre la fréquence F et la note de musique H
 - ◆ Fréquence F exprimée en Hertz (Hz) -- échelle linéaire
 - ◆ Hauteur perçue H -- échelle logarithmique



Autres paramètres : introduction



✧ Spécificités de la musique : variations de fréquence

- ◆ à l'échelle macroscopique : mélodie
- ◆ à l'échelle microscopique :

variations linéaires :



variations périodiques (< 20 Hz) :



variations périodiques plus rapides : rajout de composants spectraux



Autres paramètres : introduction

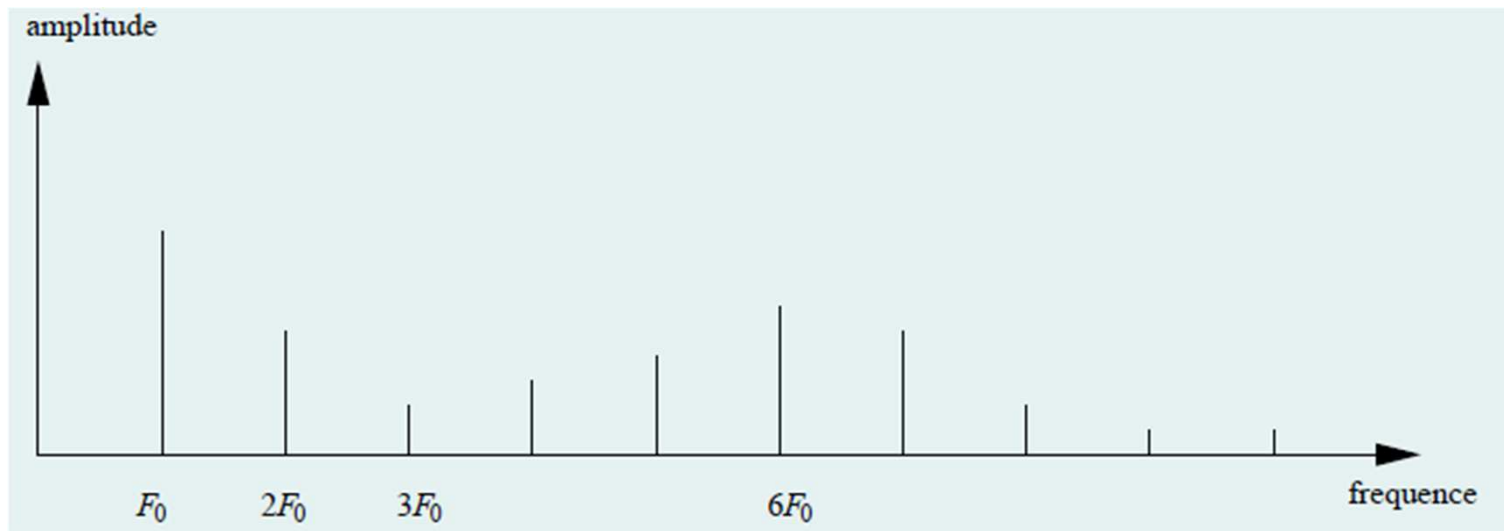


✧ Spécificités de la musique : spectre

- ◆ fréquences composant le son **régulièrement** espacées

F_0 : fréquence **fondamentale** (ou première harmonique)

$k.F_0$: $k^{\text{ème}}$ harmonique



Autres paramètres



- ✧ Introduction : spécificités de la musique
- ✧ Fréquence fondamentale
- ✧ Constant-Q Transform
- ✧ Chromas

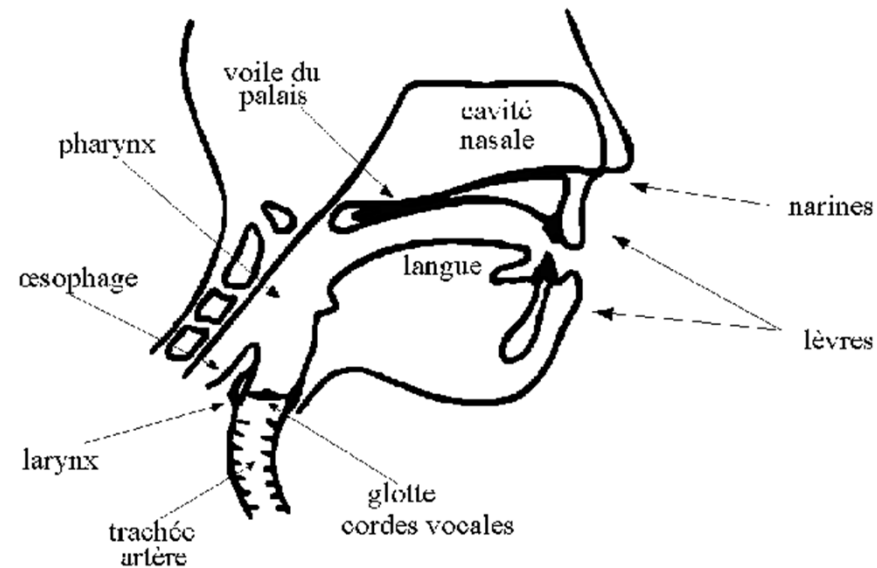


Autres paramètres : F_0



✧ Définition

- ◆ F_0 : fréquence fondamentale de vibration des cordes vocales



Autres paramètres : F_0



✧ Définition

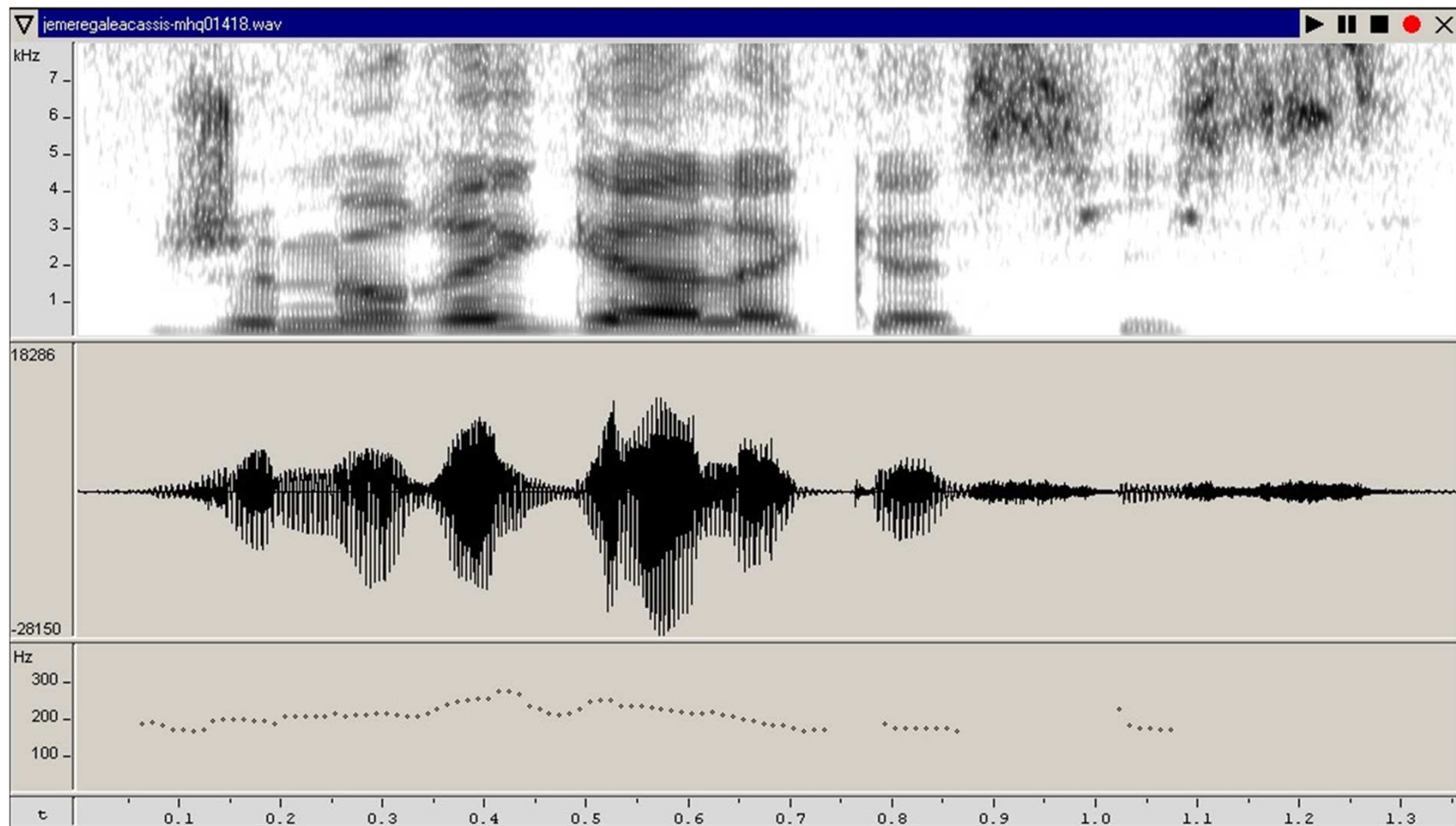
- ◆ Correspond à la hauteur de la voix (mélodie)
- ◆ Intonation
- ◆ Plage de valeur (en Hz) :
 - 80-250 :
 - 200-400 :
 - > 400 :



Autres paramètres : F_0



✧ Définition



Autres paramètres : F_0



✧ Intérêts et fonctions

- ◆ Liée à l'accentuation :
- ◆ De structuration
- ◆ Modale
- ◆ Expressive



Autres paramètres : F_0



✧ Modélisation : Fujisaki [Fujisaki 83]

- ◆ modèle de commande de source vocale
- ◆ hypothèse : les contours mélodiques sont construits à l'aide de 2 types de composantes :



Autres paramètres : F_0



✧ Modélisation : ToBI (Tones and Break Indices) [Beckman 94]

- ◆ 2 niveaux de parenthésage
- ◆ Impose une connaissance approfondie de la prosodie de la langue
- ◆ Problème : diffère d'une langue à l'autre



Autres paramètres : F_0



✧ Modélisation : INTSINT (International Transcription System for INTonation) [Hirst 91]

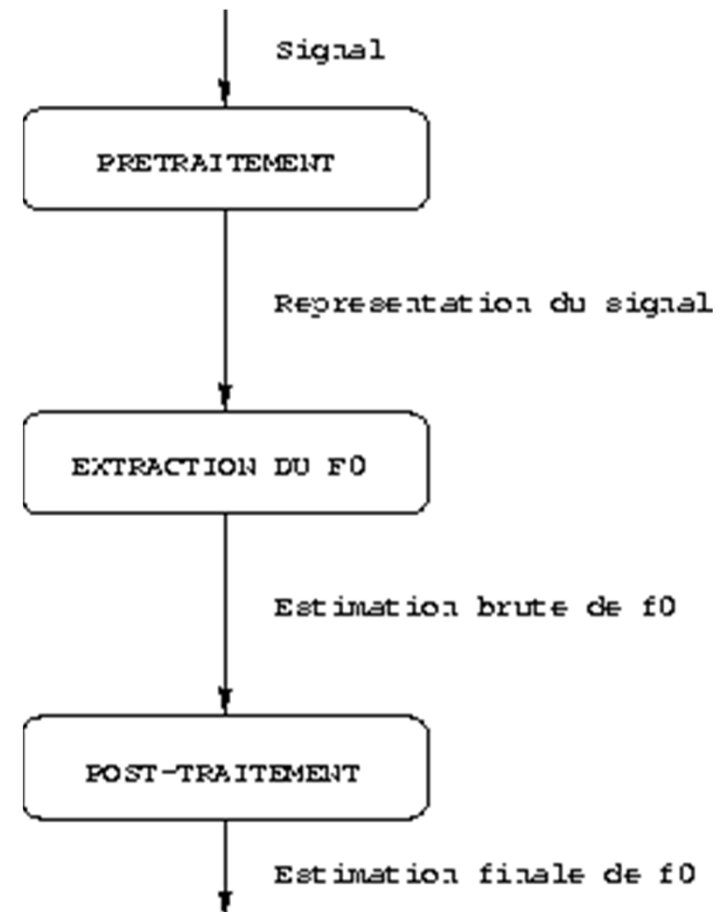
- ◆ Étiquettes absolues :
- ◆ Étiquettes relatives :



Autres paramètres : F_0



✧ Extraction : principe général



Autres paramètres : F_0



✧ Extraction

◆ Autocorrélation [Hess83]

Algorithme de type « corrélation »

Hypothèse : le signal est stationnaire



Autres paramètres : F_0



✧ Extraction

- ◆ Fonction de distance (AMDF) [Miller et Weibel 56]
critère de variation d'amplitude à court terme (Average Magnitude Difference Function)

Ambiguïté entre pics T_0 , $2 \cdot T_0$... atténuée par la non stationnarité du signal

Résiste aux erreurs grossières

Rapidité de calcul



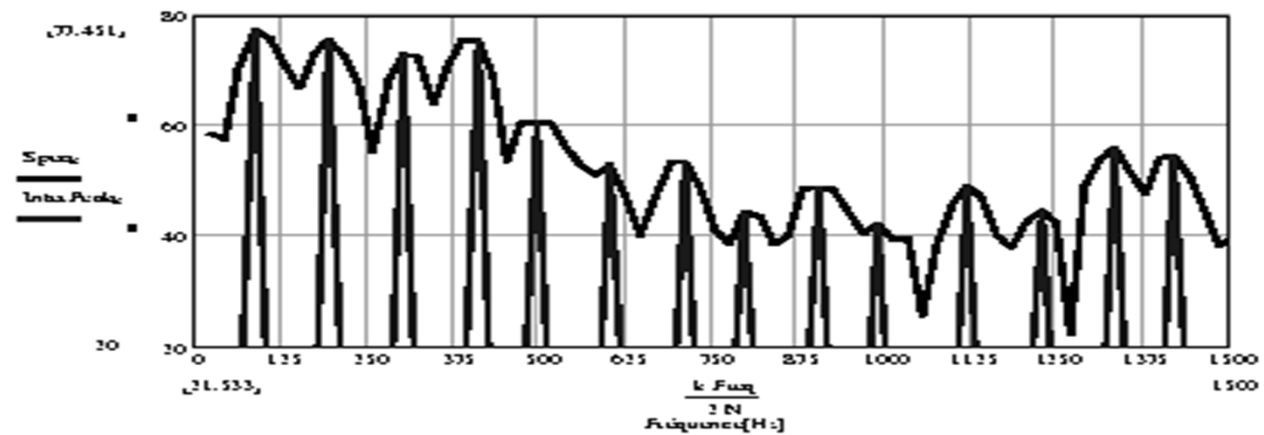
Autres paramètres : F_0



✧ Extraction

◆ Peigne spectral [Martin81]

intercorrélation entre le spectre S du signal et un « peigne »





✧ Extraction

◆ Méthodes combinatoires

combiner différentes approches pour augmenter les performances globales du système d'extraction

système de combinaison de MES SignAIX

approche basée sur la corrélation des différents candidats avec une fenêtre triangulaire [Soquet 94]





✧ Références

- ◆ [Fujisaki 83] Modelling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation, Working Group on Intonation
- ◆ [Beckman 94] Guidelines for ToBI Labelling.
- ◆ [Hirst 93] Multi-lingual modelling of intonation patterns, Workshop on prosody
- ◆ [Hess 83] Pitch determination of speech signals, Springer, 1983
- ◆ [Miller et Weibel 56] Measurements of the fundamental frequency of a speech using a delay line, JASA 28(A)
- ◆ [Martin 81] Extraction de la fréquence fondamentale par intercorrélation avec une fonction peigne, XXIIe JEP
- ◆ [Soquet 94] Approche coopérative de l'extraction de la fréquence fondamentale, XXe JEP



Autres paramètres : F_0



✧ Liens vers programmes classiques

- ◆ MES SignAix

http://www.lpl.univ-aix.fr/ext/projects/mes_signaix.htm/

- ◆ Snack/Wavesurfer

<http://www.speech.kth.se/wavesurfer/>

- ◆ Praat

<http://www.fon.hum.uva.nl/praat/>

- ◆ Modélisation automatique de la fréquence fondamentale F_0 (MOMEL)

http://www.icp.inpg.fr/~rolland/my_work/momel_french.html

- ◆ Snorri/WinSnorri

<http://www.loria.fr/~laprie/>



Autres paramètres : F_0



✧ Limite : la polyphonie (plusieurs notes jouées en même temps)

- ◆ Les méthodes précédentes sont dédiées à la monophonie
- ◆ Problème lié à la séparation de sources
- ◆ Problème très difficile
 - Sources sonores variées
 - Intervalle de notes possibles important
 - Musique : présence de batterie, de bruit
- ◆ Prise en compte du contexte tonal local



Autres paramètres



- ✧ Introduction : spécificités de la musique
- ✧ Fréquence fondamentale
- ✧ Constant-Q Transform
- ✧ Chromas



Autres paramètres : CQT



✧ Signal de musique : choix de la fenêtre d'analyse est prépondérant !

◆ Fenêtre courte (bonne résolution temporelle)

◆ Fenêtre grande





✧ Illustration

- ◆ Soit un signal de musique avec 2 sinusoides de fréquences f_1 et f_2 qui correspondent à 2 notes adjacentes
- ◆ Notons $\Delta f = f_2 - f_1$
- ◆ Pour la musique occidentale (échelle logarithmique), les fréquences correspondant à 2 notes adjacentes sont d'autant plus proches que l'on se rapproche des basses fréquences
- ◆ 2 notes adjacentes (correspondant à un demi-ton) sont séparées par 6 % de la fréquence de la note la plus basse



Autres paramètres : CQT



✧ Solution : analyse fréquentielle multi-résolution

- ◆ Résoudre le dilemme résolution temps/fréquence
- ◆ FFT (transformée en résolution fixe)
- ◆ Analyse fréquentielle multi-résolution



Autres paramètres : CQT



✧ Transformée en constante Q (Constant-Q Transform – CQT)

- ◆ Proposée par Brown en 1991
- ◆ Canaux fréquentiels pas linéairement espacés (comme FFT)
MAIS géométriquement espacés :

- ◆ Résolution temporelle augmente avec la fréquence



Autres paramètres : CQT



✧ Transformée en constante Q (Constant-Q Transform – CQT)

◆ Formule :

avec :

- $X(k)$ la $k^{\text{ième}}$ composante de la CQT,
- $x(n)$ le signal,
- $w(n, k)$ la fenêtre d'analyse

La longueur de $w(n, k)$ à la fréquence f_k est :



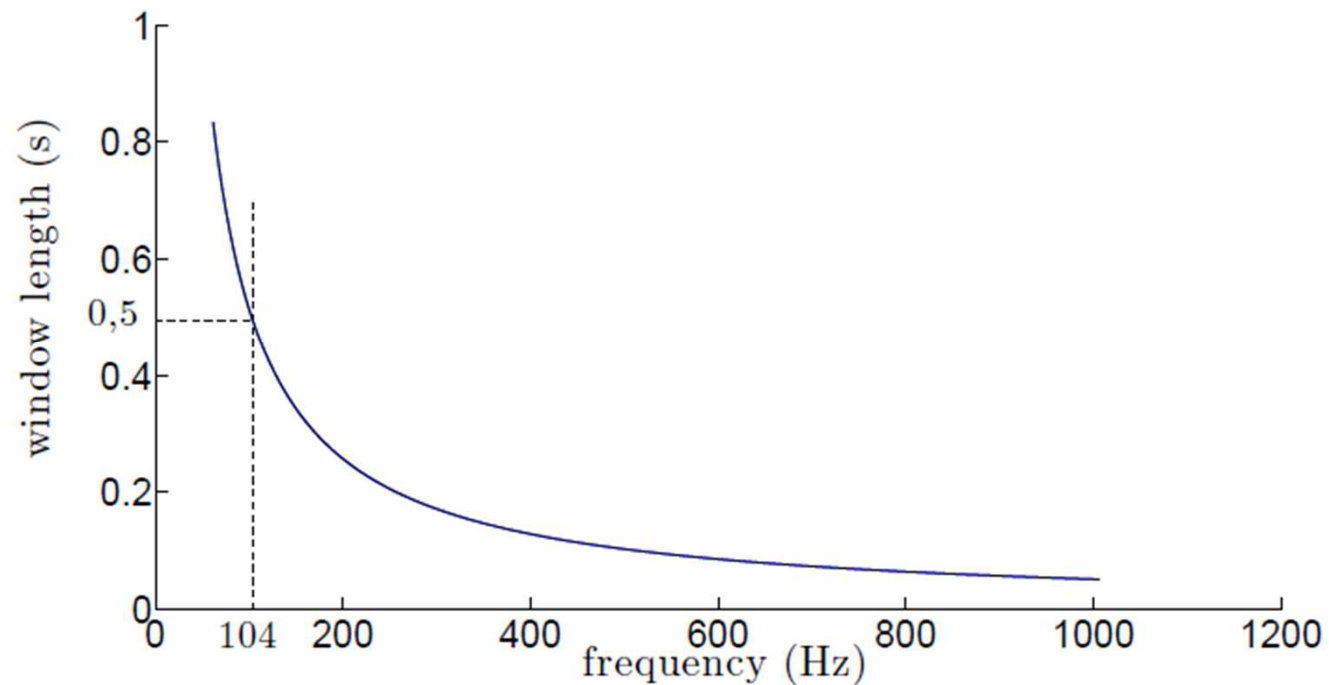
Autres paramètres : CQT



✧ Transformée en constante Q (Constant-Q Transform – CQT)

◆ Illustration

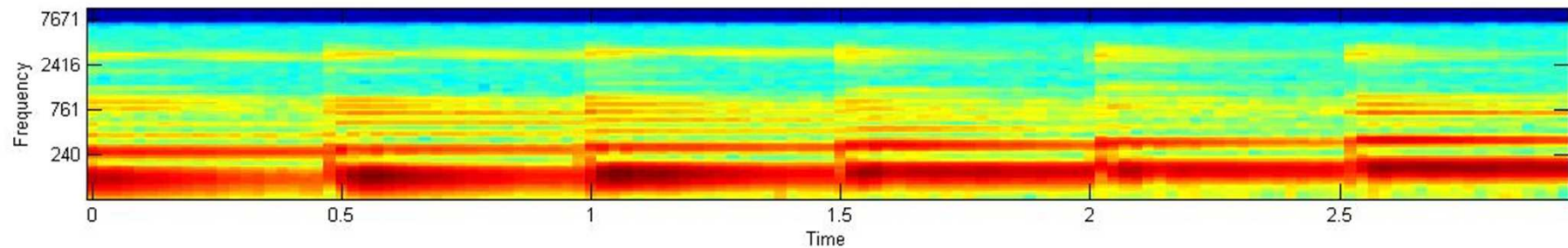
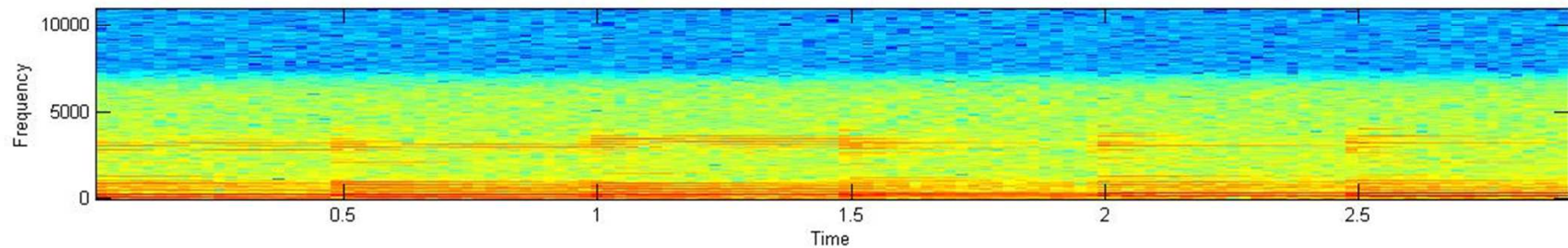
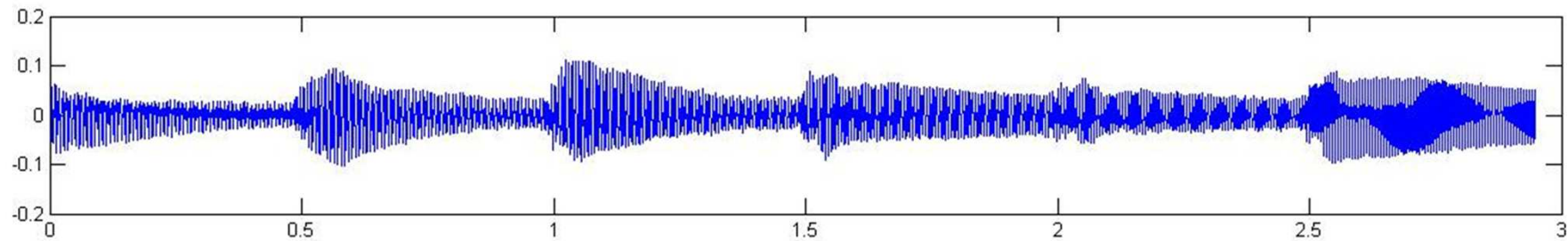
Taille de la fenêtre N (en secondes) en fonction de la fréquence (en Hertz)
pour un espacement d'un demi-ton



Autres paramètres : CQT



✧ Transformée en constante Q (Constant-Q Transform – CQT)



Autres paramètres



- ✧ Introduction : spécificités de la musique
- ✧ Fréquence fondamentale
- ✧ Constant-Q Transform
- ✧ Chromas



Autres paramètres : chromas



✧ Présentation

- ◆ Les chromas ou chroma vector ou Pitch Class Profile (PCP)
- ◆ Les paramètres « rois » pour l'analyse de la musique !

✧ Intérêt

- ◆ Représenter l'importance de chacune des 12 notes de la gamme dans un accord donné
- ◆ Il s'agit d'une représentation spectrale particulière, en dimension 12, issue de la proposition du psychologue Shepard [She64]



Autres paramètres : chromas



✧ Méthode

- ◆ Le spectre est divisé en 12 classes (12 demi-tons de la gamme)

.

- ◆ Intensité des 12 demi-tons
- ◆ Vecteur de chromas (12 dimensions)
- ◆ Normalisation possible

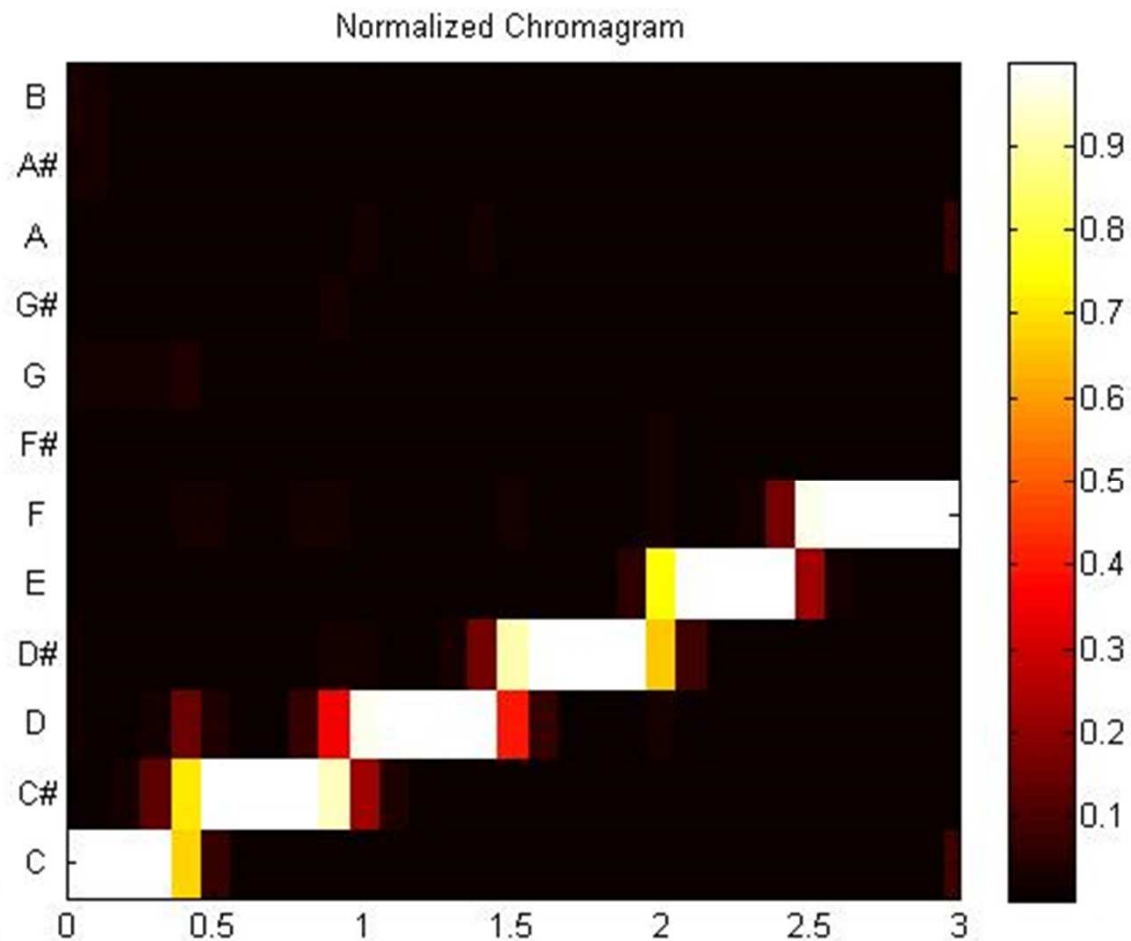


Autres paramètres : chromas



✧ Chromagram (ou spectre de chroma)

- ◆ Affichage compact de la représentation spectrale (FFT ou CQT)
- ◆ Exemple sur la gamme



Autres paramètres : chromas



- ✧ Chromagram (ou spectre de chroma)
- ◆ Exemple sur un morceau des Beatles

