

Outils Informatiques pour le Multimédia

Master 1 informatique
Tronc commun

Département d'informatique
Université Paul Sabatier

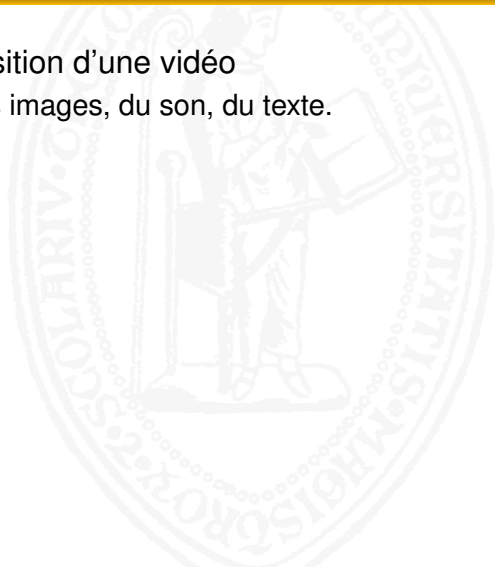
Segmentation et extraction d'informations audio/vidéo

- 1 Introduction et définitions
- 2 Segmentation temporelle
 - Vidéo
 - Audio
- 3 Extraction de caractéristiques
 - Vidéo
 - Audio

Extraction de l'information d'une vidéo

Introduction

- Composition d'une vidéo
 - Des images, du son, du texte.



Extraction de l'information d'une vidéo

Introduction

- Composition d'une vidéo
 - Des images, du son, du texte.

Extraction de l'information à partir d'une vidéo

- à partir des images
 - Couleur, texture, formes, objets ...
- à partir du mouvement
 - Traitement de séquences d'images
- à partir du son
 - Traitement de la parole et de la musique

Extraction de l'information d'une vidéo

Introduction

- Composition d'une vidéo
 - Des images, du son, du texte.

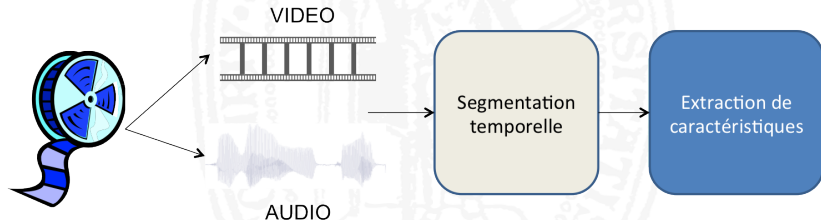
Extraction de l'information à partir d'une vidéo

- à partir des images
 - Couleur, texture, formes, objets ...
- à partir du mouvement
 - Traitement de séquences d'images
- à partir du son
 - Traitement de la parole et de la musique

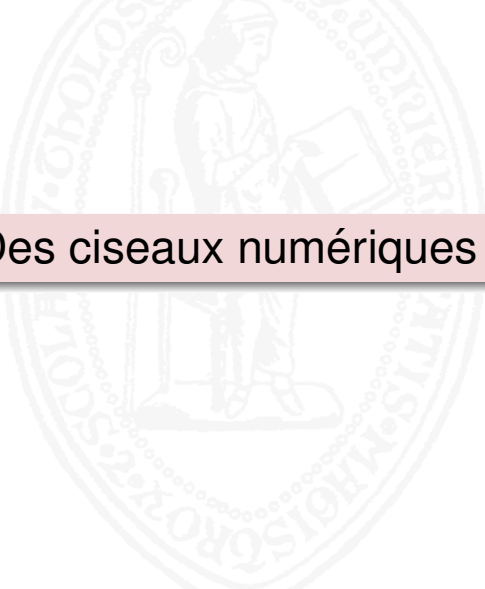
Nécessite une segmentation préalable

Extraction de l'information d'une vidéo

Segmentation d'une vidéo



Segmentation temporelle d'une vidéo

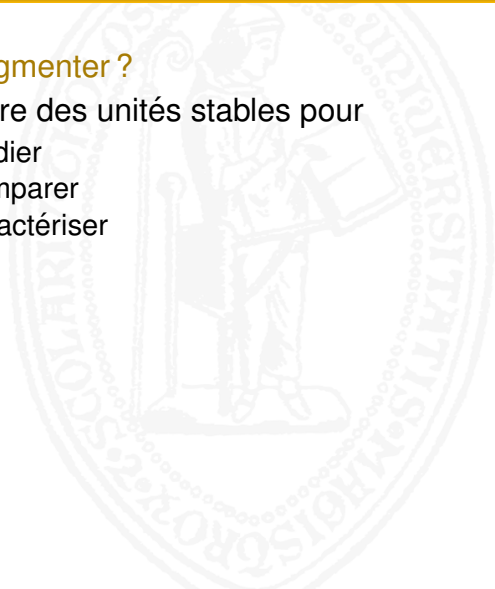


Des ciseaux numériques

Segmentation d'une vidéo

Pourquoi segmenter ?

- construire des unités stables pour
 - Etudier
 - Comparer
 - Caractériser
 - ...



Segmentation d'une vidéo

Pourquoi segmenter ?

- construire des unités stables pour
 - Etudier
 - Comparer
 - Caractériser
 - ...

Types de segmentation

- Vidéo
 - plans, scènes, chapitres, émissions, ...
- Audio
 - parole/musique/bruit, locuteurs/chanteurs, ...

Segmentation temporelle d'une vidéo

Notion de plan

- Prise de vue sans interruption
- Durée de quelques secondes
- Unité de base du langage cinématographique

Segmentation temporelle d'une vidéo

Notion de plan

- Prise de vue sans interruption
- Durée de quelques secondes
- Unité de base du langage cinématographique

Segmentation en plans

- Identification automatique des bornes des plans
- Problème le plus ancien et le plus étudié
- Brique de base pour l'analyse de la vidéo

Segmentation temporelle d'une vidéo

Segmentation en plans

- recherche des transitions dans le flux d'images
 - Franches : Coupure ou cut



Segmentation temporelle d'une vidéo

Segmentation en plans

- recherche des transitions dans le flux d'images
 - Franches : Coupure ou cut
 - Progressives : fondus, volets, ...



Segmentation temporelle d'une vidéo

Segmentation en plans

- Images fortement dissemblables au voisinage d'une transition
 - Extraction d'une caractéristique depuis les images
 - Définition d'une distance sur les caractéristiques
 - Analyse des variations sur la courbe temps/distance

Segmentation temporelle d'une vidéo

Segmentation en plans

- Images fortement dissemblables au voisinage d'une transition
 - Extraction d'une caractéristique depuis les images
 - Définition d'une distance sur les caractéristiques
 - Analyse des variations sur la courbe temps/distance

Outil de base : histogramme

Segmentation temporelle d'une vidéo

Méthodes de segmentation en plans

- Domaine décompressé
 - Comparaison de données entre les images
 - Couleurs des pixels
 - Histogrammes locaux ou globaux
 - Structures : contours, coins, ...
 - Modélisation mathématique
 - Représentation de la transition

Segmentation temporelle d'une vidéo

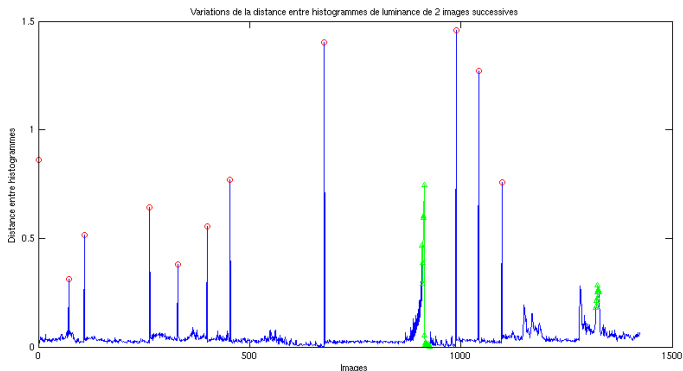
Méthodes de segmentation en plans

- Domaine compressé
 - En lien avec les propriétés du flux MPEG
 - Composante continue de la DCT
 - Analyse du mouvement
 - Analyse des blocs
 - Analyse du débit
 - ...

Segmentation temporelle d'une vidéo

Segmentation en plans : exemple

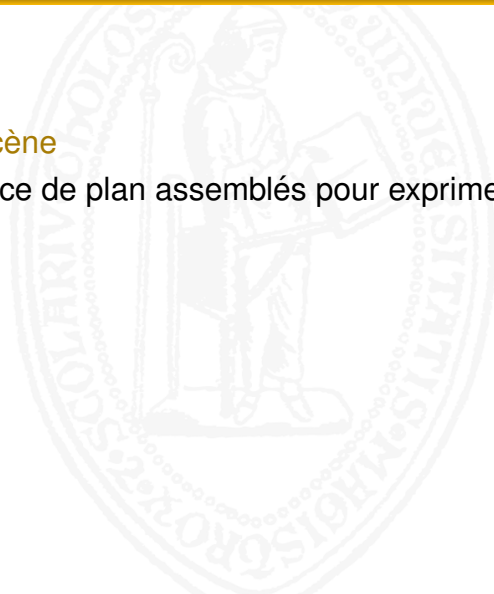
- Pour chaque image, histogramme de luminance $H(p_i)$
- Fonction de distance entre histogramme $d(h1, h2)$
- Courbe de similarité $s(i) = d(H(p_i), H(p_{i+1}))$



Segmentation temporelle d'une vidéo

Notion de scène

- Séquence de plan assemblés pour exprimer une idée



Segmentation temporelle d'une vidéo

Notion de scène

- Séquence de plan assemblés pour exprimer une idée

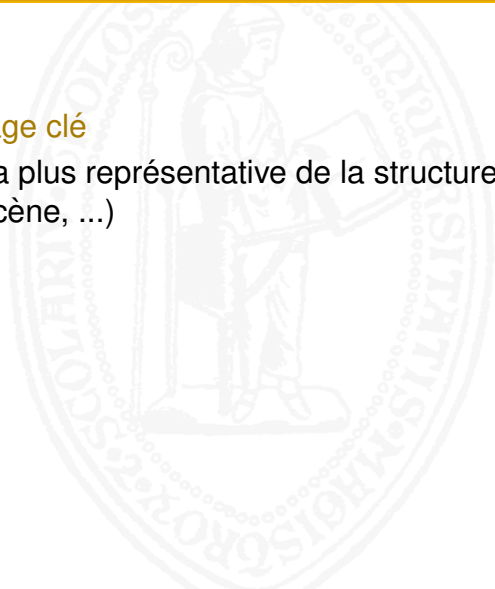
Méthodes de segmentation en scènes

- Regroupement, sous contrainte temporelle, de plans similaires
- Regroupement de plans à base de règles

Segmentation temporelle d'une vidéo

Notion d'image clé

- Image la plus représentative de la structure étudiée (plan, scène, ...)



Segmentation temporelle d'une vidéo

Notion d'image clé

- Image la plus représentative de la structure étudiée (plan, scène, ...)

Méthodes de construction d'image clé

- En théorie, représentation sémantique
- En pratique, utilisation de la couleur, de la texture, du mouvement,

Segmentation audio

Il n'y a pas que l'image dans la vidéo ...

Segmentation audio

Segmentation parole/bruit/musique

- Recherche des transitions dans le flux audio
- Notion de traits acoustiques caractéristique

Segmentation audio

Segmentation parole/bruit/musique

- Recherche des transitions dans le flux audio
- Notion de traits acoustiques caractéristique
 - pour la parole
 - pour la musique
 - pour le silence
 - pour le bruit

Segmentation audio

Segmentation parole/bruit/musique

- Traits acoustiques caractéristiques de la parole
- Liés à sa production
 - Energie du son : liée à la pression de l'air en amont du larynx
 - Son voisé : vibration des cordes vocales
 - Son non voisé (bruit) : turbulences liées à l'écoulement de l'air dans le conduit vocal
 - Durée d'un son : 60 à 100ms

Segmentation audio

Segmentation parole/bruit/musique

- Traits acoustiques caractéristiques de la musique
- Assez difficile à définir
 - produite et perçue de façons différentes

Segmentation audio

Segmentation parole/bruit/musique

- Traits acoustiques caractéristiques de la musique
- Assez difficile à définir
 - produite et perçue de façons différentes
- Pour la musique instrumentale occidentale
 - Sons harmoniques (notes)
 - Hauteur fondamentale
 - Fréquence fondamentale

Segmentation audio

Segmentation parole/bruit/musique

- Traits acoustiques caractéristiques du silence
 - En théorie, énergie nulle (0 dB)
 - En pratique, énergie seuil (40 dB)
- Traits acoustiques caractéristiques du bruit
 - Difficile à définir, tout ce qui n'est pas d'une autre classe
 - Bruits recherchés : applaudissement, rires, bruitages,

Segmentation audio

Segmentation parole/bruit/musique

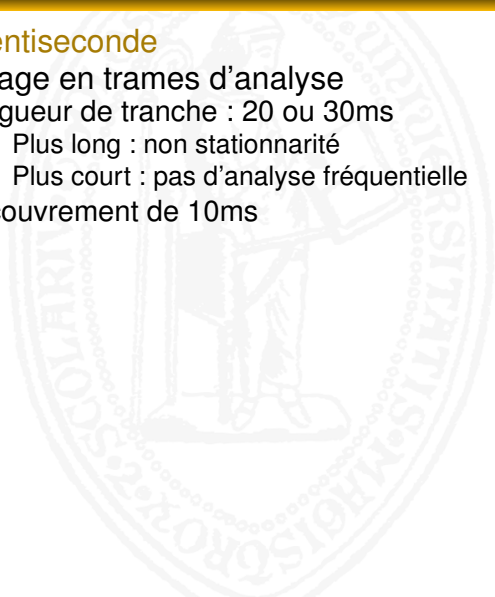
Propriétés de quasi stationnarité

- Séquence de son : succession d'états
 - Non stationnaire à long terme
 - Quasi stationnaire à court terme
 - Besoin d'une longueur perceptuelle significative
- Analyse à court terme
 - Approche centiseconde (tranches de 10ms)

Segmentation audio

Approche centiseconde

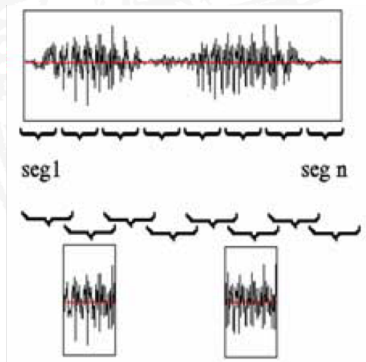
- Découpage en trames d'analyse
 - Longueur de tranche : 20 ou 30ms
 - Plus long : non stationnarité
 - Plus court : pas d'analyse fréquentielle
 - Recouvrement de 10ms



Segmentation audio

Approche centiseconde

- Découpage en trames d'analyse
 - Longueur de tranche : 20 ou 30ms
 - Plus long : non stationnarité
 - Plus court : pas d'analyse fréquentielle
 - Recouvrement de 10ms



Segmentation audio

Approche centiseconde

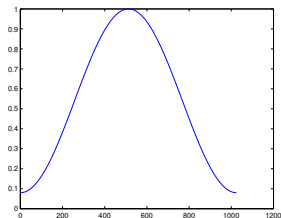
- Fenêtrage



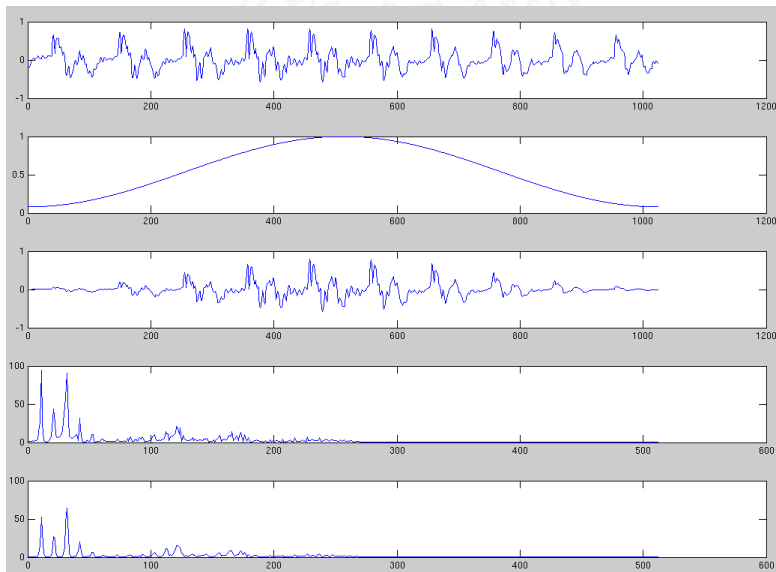
Segmentation audio

Approche centiseconde

- Fenêtrage
- Pondération des fenêtres d'Analyse
 - Réduire les effets de bord pour l'analyse fréquentielle
- Recouvrement de 10ms



Segmentation audio



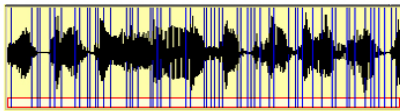
Segmentation audio

Audiogramme

- Succession de zones
 - Silence, voisées, bruitées

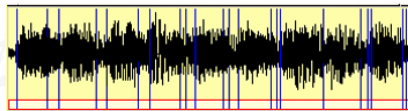
Parole Audiogramme

Signal



Musique Audiogramme

Signal



Extraction de caractéristiques

Découper n'est pas structurer ...

Extraction de caractéristiques d'une vidéo

Caractérisation et structuration de la vidéo

- Vidéo étudiée image par image
 - Couleur, texture, formes, objets, ...
 - Algorithmes d'extraction à partir d'images

Extraction de caractéristiques d'une vidéo

Caractérisation et structuration de la vidéo

- Vidéo étudiée image par image
 - Couleur, texture, formes, objets, ...
 - Algorithmes d'extraction à partir d'images
- Vidéo étudiée comme un tout
 - Caractérisation du mouvement

Extraction de caractéristiques d'une vidéo

Caractérisation du mouvement

- Mouvement de caméra
 - information riche sur l'activité d'un plan et des objets

Extraction de caractéristiques d'une vidéo

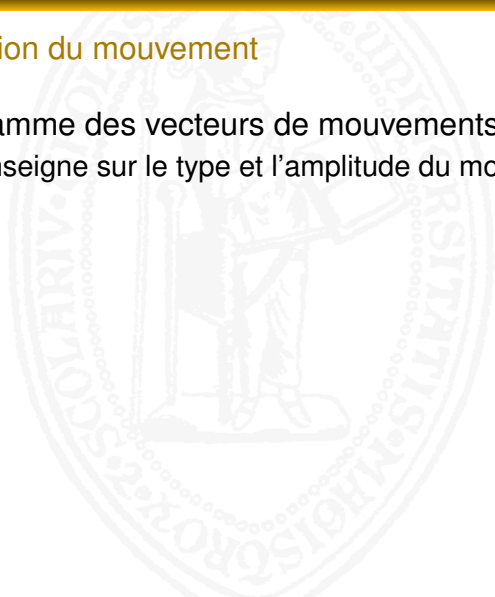
Caractérisation du mouvement

- Mouvement de caméra
 - information riche sur l'activité d'un plan et des objets
- Flux optique
 - Ensemble des vecteurs de mouvement des points (ou régions)
 - Flux fournis par formats MPEG

Extraction de caractéristiques d'une vidéo

Caractérisation du mouvement

- Histogramme des vecteurs de mouvements
 - Renseigne sur le type et l'amplitude du mouvement



Extraction de caractéristiques d'une vidéo

Caractérisation du mouvement

- Histogramme des vecteurs de mouvements
 - Renseigne sur le type et l'amplitude du mouvement
- Décomposition en ondelettes
 - Modélisation multi-résolution des vecteurs de mouvement
 - Mouvement général + détails de mouvement
- ...

Extraction de caractéristiques d'une vidéo

Caractérisation du mouvement

- Histogramme des vecteurs de mouvements
 - Renseigne sur le type et l'amplitude du mouvement
- Décomposition en ondelettes
 - Modélisation multi-résolution des vecteurs de mouvement
 - Mouvement général + détails de mouvement
- ...

Limiter les calculs par traitement des macroblocs du flux
MPEG

Extraction de caractéristiques audio

Caractérisation et structuration de l'audio



Extraction de caractéristiques audio

Caractérisation et structuration de l'audio

- Le son se définit par :
 - Volume (amplitude)
 - Vitesse de vibration (fréquence)
 - Type de voix ou d'instrument (timbre)

Extraction de caractéristiques audio

Caractérisation et structuration de l'audio

- Le son se définit par :
 - Volume (amplitude)
 - Vitesse de vibration (fréquence)
 - Type de voix ou d'instrument (timbre)
- Le codage du son se définit par :
 - La fréquence d'échantillonnage
 - Le nombre de bits de quantification

Extraction de caractéristiques audio

Exemples audio

- Fichier original de parole : 16 bits, 16 kHz
- Variation de la fréquence d'échantillonnage
 - 8 kHz
 - 4 kHz
 - 2 kHz
- Variation du nombre de bits de quantification
 - 8 bits
 - 4 bits
 - 2 bits

Extraction de caractéristiques audio

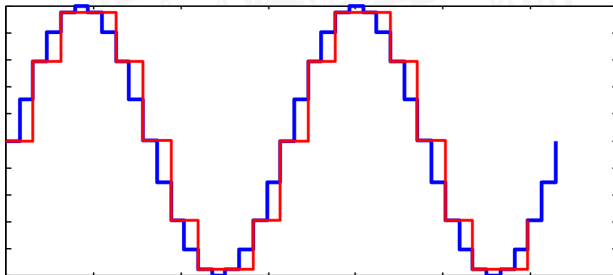
Traitements élémentaires

- Mixage
 - $\alpha \times \text{Son1} + \beta \times \text{Son2}$
- Ajouts d'effets
 - Changement de vitesse
 - Echo
 - Inversion
 - sous quantification
 - ...

Extraction de caractéristiques audio

Traitements élémentaires

- Sous échantillonnage
 - min, max, moyenne, ...



Extraction de caractéristiques audio

Extraction de paramètres temporels

- Énergie à court terme
 - Signal $x(t)$, fenêtre d'analyse de taille Δt contenant N points

$$E(t) = \frac{1}{N} \sum_{i=1}^N x_i(t)^2$$

- ZCR (Zero Crossing Rate)
 - Nombre de fois ou le signal passe à 0 sur la fenêtre d'analyse

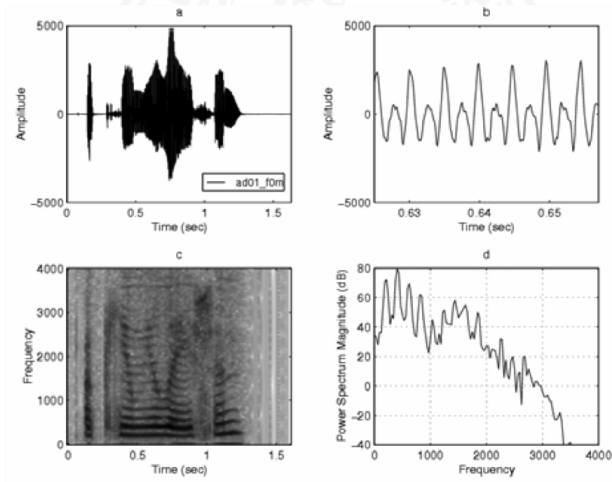
Extraction de caractéristiques audio

Extraction de paramètres fréquentiels

- Fréquence fondamentale (F_0)
 - Fréquence de vibration des cordes vocales
 - 70 à 250 Hz chez l'homme
 - 150 à 400 chez la femme
 - 200 à 600 chez l'enfant
- Formants (F_1, F_2, \dots)
 - Maxima d'énergie du spectre
- Spectrogramme
 - Représentation graphique de l'évolution du spectre
 - temps \times fréquence \times énergie

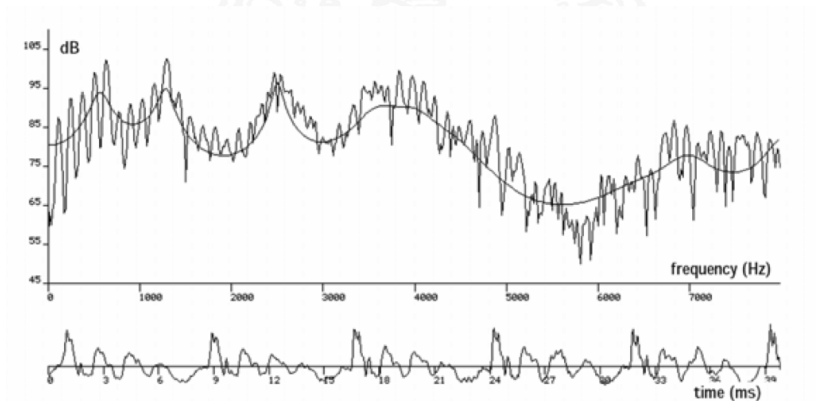
Extraction de caractéristiques audio

Exemple : analyse fréquentielle court terme



Extraction de caractéristiques audio

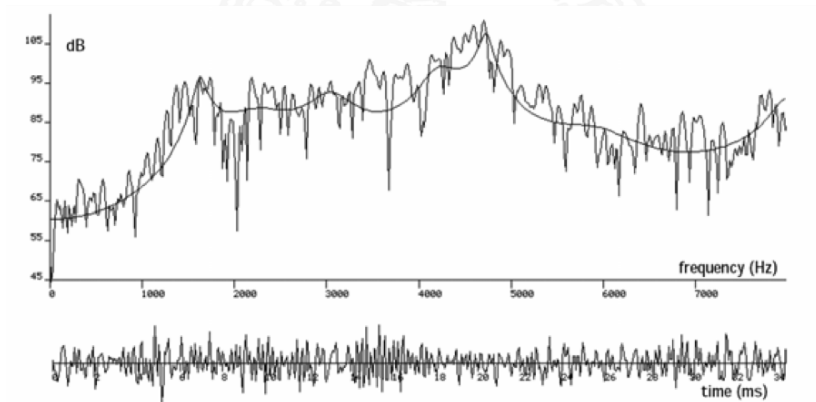
Exemple : analyse d'un son voisé



[a] de "baluchon"

Extraction de caractéristiques audio

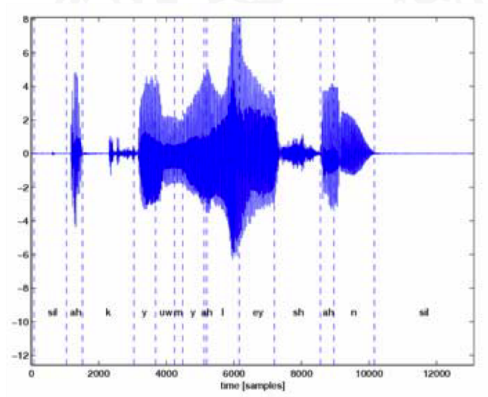
Exemple : analyse d'un son voisé



[ch] de “baluchon”

Extraction de caractéristiques audio

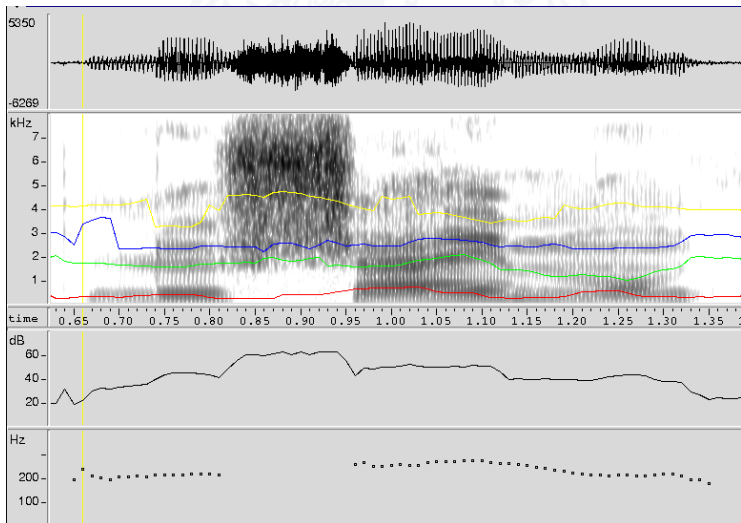
Exemple : analyse d'un son voisé



“accumulation” en anglais

Extraction de caractéristiques audio

Exemple : logiciel wavesurfer



Extraction de caractéristiques audio

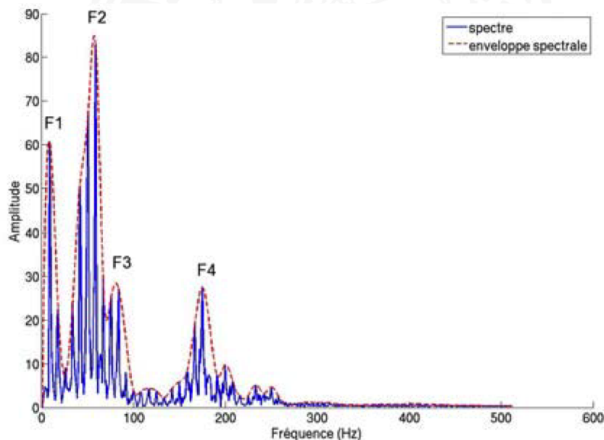
Extraction de paramètres fréquentiels

- Densité spectrale de puissance (transformée de Fourier)
 - Flux spectral
 - Centroïde spectral
 - Moments du spectre
 - les 4 premiers
 - Enveloppe spectrale

Extraction de caractéristiques audio

Enveloppe spectrale

- Parole
 - Forme générale du spectre (Sans F_0)



Extraction de caractéristiques audio

Enveloppe spectrale

- Musique
 - Forme générale du spectre

